



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Stimmungen und Qualitätsbewertung im Konferenzdolmetschen

Zum Einfluss der Stimmung auf die Evaluation von
Simultandolmetschleistungen durch Studienkolleg_innen

verfasst von / submitted by

Judyta Dyrz, BA BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 070 331 342

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Translation Deutsch Englisch

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Cornelia Zwischenberger

Für meine Eltern.

Für Pius.

Danksagung

Als erstes möchte ich mich bei Frau Univ.-Prof. Mag. Dr. Cornelia Zwischenberger für die Betreuung meiner Masterarbeit bedanken. Ich bedanke mich an dieser Stelle für all die interessanten Inputs, hilfreichen Feedbacks und konstruktiven Kommentare. Ich bin sehr dankbar für die ehrlichen Rückmeldungen und für die mir entgegengebrachte Geduld. Vielen Dank für die Begleitung und Unterstützung auf diesem spannenden Weg.

Weiters gilt mein Dank allen Lehrenden und Studienkolleg_innen, die in irgendeiner Art und Weise zum Fortschritt und Gelingen meiner Arbeit beigetragen haben. Ich danke allen, die meinen Aufruf zum empirischen Versuch geteilt und weitergeleitet haben. Danke auch an alle Kolleg_innen, die bereit waren, eine Aufnahme ihrer Verdolmetschungen zur Verfügung zu stellen und allen Kolleg_innen, die ein bisschen ihrer Zeit geopfert haben, um an meinem Versuch teilzunehmen. Das alles waren wichtige und wertvolle Schritte, und ich möchte mich herzlichst für die Hilfe und die Beiträge bedanken.

In meinem privaten Umfeld möchte ich mich besonders bei meinen Eltern bedanken, die mir in meiner akademischen Laufbahn alle Freiheiten gelassen haben und immer überzeugt waren, dass ich alles schaffen kann. Danke, dass ihr nie an mir gezweifelt habt. Und schließlich gilt ein großer Dank Pius, der mir bei allen Höhen und Tiefen in dieser Zeit zur Seite stand und mir Kraft gespendet hat. Ich möchte dir von Herzen für deine Unterstützung und jedes einzelne motivierende Wort danken.

Inhaltsverzeichnis

Abbildungsverzeichnis	9
Tabellenverzeichnis	11
1. Einleitung	13
2. Qualität.....	16
2.1. Entwicklung und Forschungsstand der Qualitätsforschung	16
2.2. Qualität als soziales Konstrukt.....	20
2.2.1 Sozialer Konstruktivismus und soziales System	21
2.2.2. Das soziale Konstrukt „Qualität“ in der Dolmetschwissenschaft	23
2.3. Evaluation von Qualität.....	30
2.3.1. Kriterien der Qualitätsbewertung	31
2.3.2. Kontext und Bewertungsperspektive	37
2.3.2.1. Die Bedeutung des situativen Kontextes.....	38
2.3.2.2. Die Rolle der Rezipient_innen	41
2.3.3. Qualitätsbewertung im didaktischen Setting.....	44
2.3.3.1. Bewertungstypologien in der Dolmetschdidaktik	45
2.3.3.2. Evaluationskriterien	49
3. Expertise.....	53
3.1. Zur Definition von Expertise.....	53
3.1.1. Erkenntnisse aus der Kognitionswissenschaft und der Begabungsforschung.....	55
3.1.2. Expertisemodelle nach Ericsson (2000) und Schneider (1992)	58
3.2. Expertise im Bereich des Dolmetschens	59
4. Affekte	64
4.1. Definitionen und Abgrenzungen	64
4.1.1. Stimmungen vs. Emotionen	67
4.1.2. Affektstruktur	70
4.2. Affekte in der Translationswissenschaft.....	74

4.3. Zur Auswirkung von Stimmungen	76
4.3.1. <i>Feelings-as-information theory</i> nach Schwarz (2011)	77
4.3.1.1. Entstehung und Entwicklung	77
4.3.1.2. Grundzüge der Theorie.....	78
4.3.2. Einfluss von Stimmungen auf die Evaluation - Studie von Martina Behr (2013a)	81
5. Methode, Versuchsdesign und Durchführung	85
5.1. Methode und Versuchsdesign.....	85
5.1.1. Ausgangspunkt und Forschungsfragen.....	88
5.1.2. Beschreibung der Versuchsteilnehmer_innen	89
5.1.3. Vorgehensweise und Durchführung	91
5.2. Methoden der Datenerhebung	94
5.2.1. Fragebogen zur Bewertung der Verdolmetschungen.....	95
5.2.2. Stimmungsfragebogen.....	96
5.3. Ergebnisse	98
5.3.1. Gesamtbenotung derselben Dolmetschung zu zwei verschiedenen Zeitpunkten...	98
5.3.2. Zusammenhang zwischen Gesamtnotengebung und Stimmung.....	100
5.3.3. Bewertung der Einzelaspekte	108
6. Diskussion	116
7. Bibliografie.....	120
Anhang	128
Abstract (Deutsch).....	132
Abstract (Englisch).....	133

Abbildungsverzeichnis

Abbildung 1: Affektgitter nach Mendelsohn et al. (1989:494)	73
Abbildung 2: Differenz bei der Benotung zwischen erstem und zweitem Termin - beide Gruppen (eigene Erhebung)	99
Abbildung 3: Stimmungsdifferenz beider Gruppen (eigene Erhebung)	101
Abbildung 4: Stimmungswerte beider Gruppen an beiden Terminen (eigene Erhebung)	102
Abbildung 5: Gegenüberstellung der Noten- und Stimmungsdifferenz für Gruppe 1 (eigene Erhebung)	103
Abbildung 6: Gegenüberstellung der Noten- und Stimmungsdifferenz für Gruppe 2 (eigene Erhebung)	103
Abbildung 7: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Gestimmtheit für Gruppe 2 (eigene Erhebung)	104
Abbildung 8: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Aktivität für Gruppe 2 (eigene Erhebung)	105
Abbildung 9: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Gestimmtheit für Gruppe 1 (eigene Erhebung)	106
Abbildung 10: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Aktivität für Gruppe 1 (eigene Erhebung)	106
Abbildung 11: Gegenüberstellung der Noten- und Stimmungsdifferenz für beide Gruppen (eigene Erhebung)	107
Abbildung 12: Gegenüberstellung der Noten- und Stimmungsdifferenz Vpn 2, 3, 4 und 5 (eigene Erhebung)	108
Abbildung 13: Differenz der Noten für die Einzelaspekte bei Gruppe 1 (eigene Erhebung)	109
Abbildung 14: Differenz der Noten für die Einzelaspekte bei Gruppe 2 (eigene Erhebung)	110
Abbildung 15: Differenz der Noten für die Einzelaspekte - beide Gruppen (eigene Erhebung)	111
Abbildung 16: Gegenüberstellung Notendifferenz für „Prosodie“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)	112
Abbildung 17: Gegenüberstellung Notendifferenz für „Lexik“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)	113
Abbildung 18: Gegenüberstellung Notendifferenz für „Angemessenheit“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)	114

Abbildung 19: Gegenüberstellung Notendifferenz für „Flüssigkeit“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung).....	114
---	-----

Tabellenverzeichnis

Tabelle 1: Qualitätskriterien nach Riccardi (2005:108ff.).....	35
Tabelle 2: Inhaltliche und kommunikativ-sprachliche Qualitätskriterien im Bewertungsbogen, eigene Aufstellung, basierend auf Behr (2013a) und ZTW (o.J.)	96
Tabelle 3: Das Übertragen der Items aus dem Affektgitter von Mendelsohn et. al (1989) nach Martina Behr (2013a:289)	97
Tabelle 4: Raster zur Auswertung der Stimmung (Gestimmtheit und Aktivität) nach Behr (2013a:291)	97
Tabelle 5: Absolute Häufigkeiten und Mittelwerte der vergebenen Noten beider Gruppen an beiden Terminen (eigene Erhebung)	100

1. Einleitung

Qualität spielte in der Geschichte des Dolmetschens immer schon eine große Rolle (Grbić 2008:232). Sie gilt als immanenter Bestandteil einer professionellen Dolmetschung und spielt daher auch in der Dolmetschwissenschaft eine zentrale Rolle (Riccardi 2005:99). Die dolmetschwissenschaftliche Auseinandersetzung mit diesem Thema begann jedoch erst in den 1980er Jahren (Gile 2003:109; Pöchhacker 2004:153). Einen zentralen Aspekt stellt hierbei die Aufstellung einer Definition von Qualität bzw. die Suche nach Qualitätsmerkmalen dar (Kalina 2002:38). Oftmals wurde daher die Frage gestellt, „how to define the distinguishing parameters of a ‘good’ and a ‘poor’ translation or interpretation“ (Grbić 2008:232). Eine umfassende Definition existiert jedoch bis heute nicht (Mack 2002:112).

Durch die soziologische Wende im Bereich des Konferenzdolmetschens verbreitet sich auch die soziologische Perspektive in der Dolmetschwissenschaft (Behr 2020:41). Folglich werden auch Aspekte wie Normen oder Rollen aufgegriffen und betrachtet (z.B. Zwischenberger 2013). So fand auch der soziale Konstruktivismus Verbreitung in der Dolmetschforschung. Von diesem ausgehend wird Qualität als relativer Begriff angesehen und als ein soziales Konstrukt verstanden, das von sozialen Systemen hervorgebracht wird (Grbić 2008; Zwischenberger 2013:17-48). Doch ohne eine umfassende Definition des Begriffs „Qualität“ oder seiner Teilaspekte können auch keine Beurteilungskriterien aufgestellt werden, die im Allgemeinen anerkannt wären (Mack 2002:112). Dabei ist „[d]ie Bewertung der Dolmetschqualität [...] zweifellos eine Schlüsselfrage in der Dolmetschwissenschaft und in der Dolmetschpraxis“ (Kutz 2005:14). Unterschiedliche Arbeiten nahmen hierfür verschiedene Parameter ins Blickfeld (Riccardi 2005:104-107).

Doch auch die angewandten Maßstäbe in einer bestimmten Situation, in der bewertet wird, sind relativer Natur. Dieser relative Charakter der Qualität bringt zudem mit sich, dass sie auch von der Person abhängt, die eine Evaluation vornimmt (Grbić 2008:240). Für unterschiedliche Nutzer_innen, die die Dienstleistung Dolmetschen in Anspruch nehmen, können so unterschiedliche Kriterien von Bedeutung sein (Gile 2003:109). So zeigten Untersuchungen, dass sich die Nutzer_innen in Hinblick auf ihre Erwartungen und Prioritäten sowie ihr Störempfinden unterscheiden (Kalina 2004:5). Behr (2013a) stellt an dieser Stelle die Rezipient_innen in der Vordergrund. Diese sind stets ein Teil der Dolmetschsituation, in die sie auch immer einen persönlichen Hintergrund mitbringen (Behr 2013a:15). Der Prozess des Verstehens bei den Zuhörer_innen sowie ihr Wille zu verstehen spielen an dieser Stelle eine zentrale Rolle. Dabei kommt beim Rezipieren auch Aspekten wie dem Interesse, den

Einstellungen und Normen oder der Befindlichkeit eine wichtige Bedeutung zu (Behr 2013a:19).

Der Aspekt der Befindlichkeit führt weiter in den Bereich der Affekte. Aus psychologischer Sicht können Affekte als Oberbegriff zu Stimmungen und Emotionen angesehen werden. Stimmungen sind in Abgrenzung zu Emotionen kontinuierlich und langfristig, richten sich auf nichts Spezifisches und haben auch keinen spezifischen Auslöser und können als relativ schwach betrachtet werden (Briner et al. 2000:15-20). Die verschiedenen Zustände von Affekten werden durch zwei bis drei Gefühlsdimensionen erfasst bzw. beschrieben. Diese gelten sowohl für Emotionen als auch Stimmungen. Ausgehend davon wurden unterschiedliche Modelle zur Erfassung der Affektstruktur formuliert (Briner et al. 2000:31-48). Welchen Einfluss Affekte bzw. Stimmungen auf die kognitiven Verarbeitungsprozesse haben können, zeigt zum Beispiel die *Feelings-as-information theory* (FAIT) nach Schwarz (2011). Dieser zufolge können Stimmungen als Informationsquelle herangezogen werden und haben somit einen Einfluss auf das Urteilen (Schwarz 2011).

Auch der Bereich der Translationswissenschaft beschäftigt sich mit Affekten, wobei sowohl die Dolmetsch- als auch die Übersetzungswissenschaft das Thema auf unterschiedliche Art untersucht und betrachtet haben (z.B. den Ouden 2017; Hubscher-Davidson 2018; Koskinen 2020). Behr (2013a) eröffnet hier einen neuen Blickwinkel, indem sie den Einfluss von Stimmungen auf die Beurteilung von Dolmetschungen durch Prüfer_innen untersucht und dabei von der FAIT nach Schwarz (2010) ausgeht. In ihrem empirischen Versuch konnte sie aufzeigen, dass vor allem die Evaluation der weniger inhaltsbezogenen Qualitätsaspekte „Funktionalität“ und „Ausdruck“ mehr Schwankungen aufwiesen als inhaltliche Aspekte. So scheinen diese Parameter gemäß der FAIT stärkeren Stimmungsschwankungen zu unterliegen (Behr 2013a:303f.). Die Theorie nach Schwarz (2011) verweist zudem darauf, dass der Stimmungseinfluss ebenfalls mit zunehmendem Grad an Expertise beim Urteilen abnimmt (Schwarz 2011:297).

Als für Behr (2013a) in ihren Ergebnissen kein klares Bild zur Relation zwischen der Notengebung und der Erfahrungszeit der Teilnehmer_innen als Prüfer_innen erkennbar war, deutete sie darauf hin, dass die Dolmetscher_innen möglicherweise schon im Zuge ihrer Ausbildung einen ausreichend hohen Grad an Expertise erlangt hätten. So verwies sie an dieser Stelle darauf, dass Untersuchungen zur Stimmungsbeeinflussung von Studienanfänger_innen und Studierenden am Ende der Ausbildung mehr Hinweise bzw. Aufschluss geben könnten (Behr 2013a:301f.).

Diese Schlussfolgerungen waren Inspiration für die Planung eines ähnlichen empirischen Versuchs mit Student_innen am Zentrum für Translationswissenschaft (ZTW) in Wien. Hierbei sollte der Einfluss der Stimmung auf die Evaluation von Dolmetschleistungen im Bereich des Simultandolmetschens bzw. Konferenzdolmetschens durch Studienkolleg_innen untersucht werden. Ausgehend von der FAIT nach Schwarz (2011) und den Erkenntnissen aus Behrs (2013a) Versuch sollte auch der Grad der Expertise miteinbezogen werden. Die vorliegende Arbeit ging somit drei Fragen nach. Zum einen sollte untersucht werden, wie sich die Bewertungen derselben Simultandolmetschung, die unwissentlich zu zwei unterschiedlichen Zeitpunkten vorgenommen wird, unterscheidet und welche Rolle dabei der Stimmung zukommt. Des Weiteren wurde die Frage gestellt, welche Unterschiede sich aufgrund des unterschiedlichen Studienfortschritts der bewertenden Personen ergeben. Schließlich sollte auch der Frage nachgegangen werden, inwieweit sich die Bewertungen der beiden nach Expertise unterteilten Gruppen in Hinblick auf die inhaltsbezogenen bzw. formalen Qualitätsparameter unterscheiden. Um diesem Forschungsinteresse empirisch nachgehen zu können, müssen zunächst die drei Konzepte Qualität, Expertise und Stimmung bzw. Affekt näher betrachtet und für die vorliegende Arbeit abgesteckt werden.

2. Qualität

„Qualität“ zählt neben Begriffen wie „Wissen“ und „Professionalisierung“ zu den wichtigsten Schlagwörtern der modernen Gesellschaft (Grbić 2008:236). „[W]e live in an era obsessed with quality control.“ (Kahane 2000:1) So ist die Frage nach der Qualität auch für den heutigen Dolmetschmarkt, der durch steigende Nachfrage und Wettbewerb geprägt ist, von großer Wichtigkeit. In der Dolmetschliteratur zählt Qualität mittlerweile zu den Themengebieten, die am häufigsten beschrieben werden (Mack 2002:110). Das Forschungsinteresse in diesem Feld spiegelt sich auch in den Konferenzen wider, die das Thema aufgreifen. Die Konferenz *Quality in Conference Interpreting* in Almunecar im Jahr 2001 war die erste internationale Konferenz, die ausschließlich diesem Thema gewidmet war (Kurz 2003:3).

Im Vordergrund der Dolmetschwissenschaft steht die Aufstellung einer Definition von Qualität bzw. von Qualitätsmerkmalen, sowie die Frage, wie man diese in weiterer Folge messen kann (Kalina 2002:38). Ein Blick in die Literatur zur Dolmetschqualität der letzten Jahrzehnte macht jedoch deutlich, dass sowohl verschiedene Vorstellungen von als auch unterschiedliche methodische Zugänge zu Qualität existieren (Zwischenberger 2013:17). Sie zu untersuchen, bedeutet, verschiedene Blickwinkel und Perspektiven einzunehmen und Dolmetschen sowohl als Produkt als auch Dienstleistung zu untersuchen (Kurz 2003:3). Bis heute konnte jedoch noch keine umfassende Definition von „Qualität“ oder ihrer Teilaspekte formuliert werden (Mack 2002:112).

Im vorliegenden Kapitel soll zunächst ein kurzer Einblick in die Entwicklung und den Forschungsstand der Qualitätsforschung im Konferenzdolmetschen gegeben werden, wobei einzelne Aspekte in den nachfolgenden Kapiteln detaillierter aufgegriffen werden. Nach dem Überblick wird auf eine vom sozialen Konstruktivismus geprägte Sicht auf Qualität, sprich der Auffassung von Qualität als soziales Konstrukt (Grbić 2008; Zwischenberger 2013:17-48.) eingegangen. Dem folgt die Auseinandersetzung mit der Evaluation von Qualität.

2.1. Entwicklung und Forschungsstand der Qualitätsforschung

Mit der Professionalisierung des Berufs Dolmetscher_in ab den 1950er Jahren entstanden auch erste Abhandlungen zu dem Thema. Diese waren jedoch eher intuitive Erfahrungsberichte von praktizierenden Dolmetscher_innen (Behr 2013a:24). Zu dieser Zeit war Qualität nicht explizit im Fokus des Forschungsinteresses, sondern galt implizit als Bestandteil der professionellen

Tätigkeit (Pöchhacker 2013:37). Dieser professionelle Status umfasste Kriterien wie eine universitäre Ausbildung und die als professioneller Standard geltenden Verhaltensregeln. Ein bedeutender Meilenstein wurde 1953 durch die Gründung der AIIC, des internationalen Verbandes für Konferenzdolmetscher_innen, erreicht (Pöchhacker 2012:20).

Zwischen 1965 und Mitte der 1970er Jahre begannen zunächst Forscher_innen aus anderen Disziplinen mit der Untersuchung einzelner Aspekte des Dolmetschens (Behr 2013a:24). Erste wissenschaftliche Untersuchungen zur Dolmetschqualität wurden von Forscher_innen im Bereich der Psychologie durchgeführt. Hierbei wurden die Äquivalenz zwischen Ausgangstext (AT) und Zieltext (ZT) untersucht und Fehler analysiert. Solche Untersuchungen, in denen das Original mit der Verdolmetschung verglichen wurde, wurden jedoch meistens in experimentellen Situationen und zudem mit nicht-professionellen Dolmetscher_innen durchgeführt. Die Ausgangstexte wurden vorgelesen und für die Analyse wurden anschließend Transkripte der Verdolmetschungen angefertigt. Demnach waren es in diesen Fällen die Merkmale auf der Mikroebene, die die Qualität definierten (Kalina 2004:3). Da diese frühen Untersuchungen auf experimentell gewonnenen Daten und der Analyse von Transkriptionen basierten, waren sie praxisfern (Mack 2002:112). Denn „[d]as Produkt der Verdolmetschung im Sinne einer Transkription ist [...] immer ein Kunsterzeugnis, kein hier und jetzt im Zusammenspiel der Kommunikationspartner konstruiertes Sprechhandeln mehr“ (Mack 2002:112). Aufgrund der Heranziehung von Kriterien quantitativer Art wurden Aspekte wie bewusste Auslassungen oder Zusammenfassungen meistens negativ gewertet (Mack 2002:112).

An dieser Stelle ist die Arbeit von Barik (1971) zu nennen, die sich mit der Äquivalenzbeziehung von AT und ZT beschäftigte und eine wichtige Rolle für die weitere Auseinandersetzung mit Qualität in der Wissenschaft spielte (Behr 2013a:2). In seiner Untersuchung beschreibt Barik (1971) drei Kategorien, nämlich *omissions*, *additions* sowie *substitutions and errors*, als die drei Möglichkeiten, wie der ZT vom AT abweichen kann (Barik 1971:199-207). Trotz einiger Mängel, denen sich Barik auch bewusst war, schaffte er eine wichtige Basis für die weitere Forschung auf dem Gebiet der Qualitätsforschung im Dolmetschen (Behr 2013a:28f.).

In den 1980er bzw. 1990er Jahren wurde erkannt, welche Bedeutung Aspekte wie die Situation, der Kontext und die Arbeitsbedingungen in Hinblick auf die Qualität beim Dolmetschen haben (Kalina 2004:4). Auch die Ansichten der beteiligten Personen rückten ins Zentrum des Forschungsinteresses. So wurden Kommunikationsteilnehmer_innen und Auftraggeber_innen sowie Lehrende zu ihren Ansichten befragt. Ziel war es, zum einen

abstrakte Kriterien für Qualität zu formulieren, und zum anderen konkrete Dolmetschleistungen zu beurteilen (Mack 2002:112). An dieser Stelle wurden vor allem Befragungen von Rezipient_innen durchgeführt. Dabei fand man heraus, dass die Situation sowie das Kommunikationsereignis einen Einfluss auf die wahrgenommene Qualität haben (Riccardi 2005:104f.). So befasste sich der Großteil dieser Studien damit, was als Qualität erwartet bzw. wahrgenommen wird. Aber auch Studien, in denen Qualität durch den Vergleich von AT und ZT tatsächlich bewertet wurde, wurden durchgeführt (Pradas Macías & Zwischenberger 2022:244).

Dolmetschen wurde nun empirisch und interdisziplinär erforscht und wurde als ein komplexer Prozess angesehen (Behr 2013a:25). Als Pionierarbeit im Bereich der Dolmetschqualität gilt die Arbeit von Hildegund Bühler (Zwischenberger 2013:43). Sie wandte Erkenntnisse aus dem Bereich der Textlinguistik auf das Dolmetschen an und trug zur Differenzierung des Begriffes Dolmetschqualität bei. Die von ihr aufgestellten Qualitätskriterien wurden in vielen folgenden Arbeiten aufgenommen (Kalina 2004:4f.). Sie war die erste, die die relative Wichtigkeit von einzelnen Parametern explizit und auf systematische Art und Weise untersuchte (Pöchhacker 2012:21). Ihre Arbeit gilt auch als die erste empirische Studie zu Erwartungen in diesem Bereich (Pradas Macías & Zwischenberger 2022:244).

Die Sicht der Rezipient_innen auf Qualität wurde zum zentralen Thema, was zu Befragungen verschiedener Rezipient_innen-Gruppen in unterschiedlichen Situationen führte. Dies sollte den Dolmetscher_innen eine Ausrichtung auf die Anforderungen und Erwartungen der Rezipient_innen ermöglichen. Es zeigte sich jedoch schnell, dass die Erwartungen, sowie die Prioritäten der Rezipient_innen und ihr Störempfinden recht heterogen ausfallen (Kalina 2004:5). Mit der Zeit stiegen die Erwartungen der User_innen an die Dolmetschqualität auch dramatisch an (Kurz 2003:4). Durch den Fokus auf die Perspektive der Kund_innen fanden nun auch Kriterien Beachtung, die sich ausschließlich auf den Zieltext beziehen. Bis heute wird in der Forschung zwischen inhaltlichen und formalen Parametern unterschieden (Behr 2013a:32f.). Bis zum Ende des 20. Jahrhunderts konzentrierte sich die Forschung vor allem darauf, Evaluationskriterien aufzustellen und die Frage zu beantworten, von wem Beurteilungen durchgeführt werden sollten (Pradas Macías & Zwischenberger 2022:243).

Mittlerweile gilt Dolmetschen in der Qualitätsforschung als ein komplexer und dynamischer, aus mehreren Phasen bestehender Prozess, der aus einer holistischen Perspektive betrachtet werden muss (Behr 2013a:86). So soll „eine Betrachtung des Dolmetschens im Gesamtgefüge aller damit verbundenen Faktoren und Abläufe“ (Behr 2013a:38) garantiert

werden. Solche Betrachtungsweisen sind beispielsweise in Arbeiten zu Situationalität und den Konferenztypen, sowie zu den unterschiedlichen Anforderungen an den komplexen, in Schritte unterteilten Prozess des Dolmetschens, zu finden (Behr 2013a:38-45).

So beschreibt beispielsweise Pöchhacker (1994) die Konferenz als Hypertext. Dieser wiederum umfasst weitere individuelle Texte (Pradas Macías & Zwischenberger 2022:249). Pöchhacker (1994:51) unterscheidet sieben Arten von Konferenzen für das Simultandolmetschen und definiert den Hypertext als „die Handlungen einzelner Textproduzenten und -rezipienten“ (Pöchhacker 1994:48). Statt nach allgemein gültigen Qualitätsdefinitionen oder einzelnen Kriterien zu suchen, stellt er das Dolmetschen als komplexes Handlungsgefüge sowie die Situation, die als variierender Faktor gesehen wird, in den Mittelpunkt (Behr 2013a:40f.).

Kalina (2002:40) sieht es beispielsweise als Aufgabe der Dolmetschforschung, eine erweiterte Perspektive einzunehmen, über die zwei Phasen der Rezeption und Produktion eines Textes hinauszugehen und den Dolmetschprozess „als umfassenderen Kommunikationsakt mit einer Vielzahl von Beteiligten, mit unterschiedlichen Erwartungen, Zielsetzungen und Bedürfnissen zu betrachten und dabei auch Faktoren einzubeziehen, die dem eigentlichen Dolmetschprozess vor- bzw. nachgelagert sind“ (Kalina 2002:40). Kalina (2004) geht davon aus, dass alle Beteiligten zusammenarbeiten müssen, damit eine gute Dolmetschqualität möglich ist. Sie schlägt ein erweitertes Anforderungsprofil vor, das die unterschiedlichen Verantwortungen der Akteur_innen und unterschiedliche Kompetenzen in zeitlich aufeinanderfolgenden Phasen berücksichtigt (Kalina 2004:6f.). Dieses soll zur Qualitätssicherung im Bereich Dolmetschen beitragen (Kalina 2004:2).

Behr (2013b:86) fasst den aktuellen Erkenntnisstand in drei Ebenen zusammen, die laut ihr allesamt berechtigt sind und für die Erforschung der Qualität berücksichtigt werden müssen. Zum einen geht es um die Äquivalenz von AT und ZT. Zweitens muss die Perspektive, aus der eine Evaluation vorgenommen wird, Berücksichtigung finden. Schließlich muss im Sinne der holistischen Herangehensweise der gesamten Dynamik eines Dolmetschprozesses in seiner Komplexität Rechnung getragen werden. So sind bei der Frage nach Qualität auch die Situation, die die Auswahl und Gewichtung der heranzuziehenden Kriterien bestimmt, zu berücksichtigen. Für die Qualität ist somit der Aspekt der Wirkungsäquivalenz bei Rezipient_innen in der jeweiligen Situation von Bedeutung (Behr 2013b:86). Schon Déjean Le Féal (1990:155) schreibt hierzu:

What our listeners receive through their earphones should produce the same effect on them as the original speech does on the speaker's audience. It should have the same cognitive content and be presented with equal clarity and precision in the same type of language. (Déjean Le Féal 1990:155)

Hierzu schlägt beispielsweise Pöchhacker (2012) einen Ansatz vor, den er als funktional-kognitiv¹ beschreibt und nach dem die Messung der Wirkungsäquivalenz im Fokus steht. Da es zu den Aufgaben der Dolmetscher_innen gehört, die von den Sprecher_innen erwünschten Wirkungen auf die Rezipient_innen zu erzielen, muss in der Qualitätsforschung empirisch untersucht werden, inwiefern Dolmetschungen es den Rezipient_innen ermöglichen, den Redebeitrag auch in den Zielsprachen äquivalent zu verstehen (Pöchhacker 2012:20,23ff.).

Konferenzdolmetschen wurde erst zu Beginn des 21. Jahrhunderts aus einer soziologischen Perspektive erforscht (Pradas Macías & Zwischenberger 2022:248). Behr (2020) verortet den *sociological turn* in den 2010er Jahren, in denen die soziologische Perspektive Verbreitung findet (Behr 2020:41). Prunč (³2012) spricht von einer Translationssoziologie und meint, sie

stellt ein fruchtbares interdisziplinäres Forschungsfeld dar, auf welchem in der Translationswissenschaft implizit angesprochene Sachverhalte in einem methodisch konsistenten Rahmen neu beleuchtet werden können. Durch die Zuwendung der Wissenschaft zur konkreten sozialen Bedingtheit der Berufsfelder [...] könnte die Kluft zwischen Theorie und Praxis [...] überbrückt werden. (Prunč ³2012:335)

So fand auch die Perspektive des sozialen Konstruktivismus Eingang in die Dolmetschwissenschaft. So wird Qualität nun als etwas Relatives betrachtet und als ein soziales Konstrukt beschrieben, das soziale Systeme hervorgebracht haben (Grbić 2008; Zwischenberger 2013:17-48). Die vorliegende Arbeit stützt sich auf diese Sicht der Qualität als soziales Konstrukt. So soll nun ein näherer Blick auf diese Sichtweise geworfen werden.

2.2. Qualität als soziales Konstrukt

Mittlerweile herrscht in der Dolmetschwissenschaft Konsens über die Schwierigkeit, eine einheitliche und umfassende Definition von Qualität aufzustellen. Paradoxerweise wurde dieser Versuch mit zunehmender wissenschaftlicher Tätigkeit, die eine komplexere und weiter gefasste Sichtweise mit sich brachte, auch immer schwieriger (Garzone 2003:23). Stattdessen

¹ Zum Projekt „QuaSI“-Experimente, in denen dieser Ansatz zur Anwendung kommt, siehe z.B. Pöchhacker (2012).

wird Qualität als ein relatives Konzept betrachtet (Grbić 2008:232; Harvey & Green 2000:17; Pöchhacker 2004:153; Riccardi 2003:257; Viezzi 2007:172) und als soziales Konstrukt angesehen (Grbić 2008; Zwischenberger 2013). Die Relativität von Qualität bezieht sich auf die Verwendung des Begriffs, wobei diese sowohl von den Personen als auch Situationen abhängt (Harvey & Green 2000:17). Qualität ist einer Sache nicht inhärent, sondern wird einem Objekt zugeschrieben und ist somit subjektiver Natur (Grbić 2008:234). Bevor jedoch auf das soziale Konstrukt der Qualität in der Dolmetschwissenschaft eingegangen wird, sollen zunächst der Sozialkonstruktivismus und das soziale System näher betrachtet werden.

2.2.1 Sozialer Konstruktivismus und soziales System

Beim sogenannten sozialen Konstruktivismus handelt es sich um einen methodischen Zugang, welcher von verschiedenen Disziplinen, darunter der Soziologie, geprägt wurde (Zwischenberger 2013:19). In den soziologischen Paradigmen stellt der sogenannte Sozialkonstruktivismus die früheste Form des Konstruktivismus dar und beschäftigt sich damit, dass Menschen die Wirklichkeit als Tatsache betrachten, und nicht als etwas, das von ihnen geschaffen wurde. Dies führen Berger und Luckmann (1969) in ihrer Arbeit weiter (Richter ²2016:202f.). Sie gelten als die Wegbereiter des Sozialkonstruktivismus und stellen die These auf, dass das Generieren und die Erhaltung sozialer Phänomene durch soziale Praktiken der Menschen erfolgen (Zwischenberger 2013:19). Gesellschaft wird als ständiger, dialektischer Prozess angesehen, der die drei Komponenten Externalisierung und Objektivierung, sowie Internalisierung umfasst. Diese laufen nicht etwa in zeitlicher Folge nacheinander ab, sondern charakterisieren simultan die Gesellschaft und ihre Teile (Berger & Luckmann ²⁵2013:139). Zwischenberger (2013) fasst die drei Prozesse folgendermaßen zusammen: „Menschen externalisieren, indem sie beispielsweise [...] eine Geschichte erzählen. Diese Geschichte wird dann von anderen erneut erzählt und wird im Laufe der Zeit zu einem Faktum, welches wiederum von anderen als Wahrheit internalisiert wird.“ (Zwischenberger 2013:19)

Wie bereits angedeutet, gibt es also nicht die *eine* Definition bzw. Beschreibung, die den sozialen Konstruktivismus eindeutig klassifiziert (Burr 1995:2; Zwischenberger 2013:19). Dennoch können einige Kennzeichen, die den sozialkonstruktivistischen Ansätzen gemein sind, festgemacht werden. Zunächst hält uns der soziale Konstruktivismus dazu an, unsere Auffassung der Welt zu hinterfragen und diese nicht als objektiv gegeben zu sehen (Burr 1995:2). Denn „just because we think of some music as ‘classical’ and some as ‘pop’ does not

mean we should assume that there is anything in the nature of the music itself that means it has to be divided up in that particular way” (Burr 1995:2). All die Sichtweisen, Kategorien und Konzepte, auf denen unser Verstehen der Welt basiert, sind zudem historisch und kulturell bedingt und somit relativ. Sie sind Produkte einer Kultur und Zeitepoche und hängen u.a. von ihren sozialen Gegebenheiten ab. Unser Wissen von der Welt wird schließlich durch soziale Prozesse konstruiert, ausgehandelt und erhalten (Burr 1995:3). Unsere Wahrheit, sprich „what we regard as ‘truth’ [...] our current accepted ways of understanding the world, is a product not of objective observation of the world, but of the social processes and interactions in which people are constantly engaged with each other” (Burr 1995:3).

Somit ist ein zentraler Aspekt die Ablehnung der Idee von Unvermeidbarkeit (Grbić 2008:234; Zwischenberger 2013:20). Der soziale Konstruktivismus nimmt in Hinblick auf den Status quo eine kritische Stellung ein (Hacking ³2002:19) und meint: „X hätte nicht existieren müssen oder müßte [sic] keineswegs so sein, wie es ist. X - oder X, wie es gegenwärtig ist - ist nicht vom Wesen der Dinge bestimmt; es ist nicht unvermeidlich.“ (Hacking ³2002:19) Es sind die sozialen Kräfte, die X geprägt haben (Hacking ³2002:20). Dies heißt wiederum nicht, dass das zu erforschende Objekt, das als soziales Konstrukt aufgefasst wird, nicht existiert. Denn zu behaupten, dass beispielsweise „die gehörlose Person“ ein soziales Konstrukt darstellt, bedeutet nicht, dass die Person ein Konstrukt ist, sondern die Kategorisierung bzw. Idee sozial konstruiert ist (Grbić 2008:234).

Der Ansatz des Konstruktivismus ist eng mit jenem der sozialen Systeme verknüpft (Zwischenberger 2013:21). Ein soziales System kann nach Hejl (⁷2003) als eine Gruppe von lebenden Systemen definiert werden (Hejl ⁷2003:127), wobei die Mitglieder der Gruppe „eine gemeinsame Realität und damit einen Bereich sinnvollen Handelns und Kommunizierens erzeugt haben *und* auf ihn bezogen interagieren“ müssen (Hejl ⁷2003:128). Nur wenn es den Mitgliedern möglich ist, an der Realitätskonstruktion mitzuwirken, handelt es sich um ein soziales System. Seine Grenzen können anhand der Interaktionen der einzelnen Systemkomponenten verstanden werden. Ein Individuum ist jedoch Teil vieler sozialer Systeme in einer Gesellschaft und konstituiert diese mit. Auch wenn es sich für die Teilnahme an einem sozialen System entscheidet, was ihm freigestellt ist, behaltet es jedoch seinen Charakter als Individuum. Es kann als Schnittpunkt bzw. Knotenpunkt verschiedener Systeme gesehen werden, die wiederum ein Netzwerk an Systemen, sprich die Gesellschaft, darstellen. Zudem ist die Umwelt eines sozialen Systems seinen Mitgliedern immer direkt zugänglich und die systemrelevante Realität stellt nicht die einzige Realität dar. Da die Mitglieder eines sozialen

Systems eine gemeinsame Realität erzeugen, die wiederum als Bedingung für ähnliche Interaktionen gilt, können sie als „synreferentiell“ bezeichnet werden (Hejl ⁷2003:128-136).

2.2.2. Das soziale Konstrukt „Qualität“ in der Dolmetschwissenschaft

In der Dolmetschwissenschaft fassen Grbić (2008) und Zwischenberger (2013) Qualität explizit als soziale Konstrukte auf. Sie beziehen sich dabei auf den sozialen Konstruktivismus und die oben beschriebene Definition von sozialen Systemen nach Hejl (⁷2003). Von dieser ausgehend können für das Dolmetschen die Bereiche Forschung, Ausbildung und professionelle Praxis als Systeme angesehen werden (Grbić 2008:235). Sie alle entwickeln ihre spezifischen Sichtweisen auf Qualität (Harvey & Green 2000:17). Denn der Qualitätsdiskurs findet nicht im Vakuum statt, sondern wird von dynamischen und unvorhersehbaren sozialen Prozessen initiiert (Grbić 2008:251). Hierbei handelt es sich „jedoch nicht um verschiedene Perspektiven auf ein und dieselbe Sache, sondern um verschiedene Perspektiven auf unterschiedliche Dinge, die allerdings mit dem gleichen Begriff belegt werden“ (Harvey & Green 2000:17). So schaffen sich die Mitglieder dieser Systeme eine gemeinsame Realität, wobei auch unvereinbare Realitäten innerhalb eines Systems existieren. Da ein Individuum Teil mehrerer Systeme sein kann, können beispielsweise Forscher_innen auch Lehrende sein, oder professionelle Dolmetscher_innen auch Forschung betreiben. Die einzelnen Systeme sind also nicht stabil, sondern werden von außen, sprich von anderen Systemen, beeinflusst und zeichnen sich daher sowohl durch Konservatismus als auch sozialen Wandel aus (Grbić 2008:235f.).

Betrachtet man Qualität als soziales Konstrukt, rücken auch soziale Normen und Rollen² in den Fokus (Zwischenberger 2013). Die vorherrschenden Normen beziehen sich hierbei auf akzeptierte und weitergegebene Regeln für ein bestimmtes gesellschaftliches Verhalten, das innerhalb einer Gruppe weitergegeben wird (Mack 2002:113). Sie sind mit Qualität verbunden „via the roles the interpreter has to play according to the expectations held by others and the interpreters themselves“ (Pradas Macías & Zwischenberger 2022:248). Soziale Rollen sind somit „a bundle of *normative* behavioural *expectations* that are directed at the occupants of social positions“ (Pradas Macías & Zwischenberger 2022:249). Rollen selbst können als Konstrukt betrachtet werden und dienen der Untersuchung, Ergründung und Erklärung des Verhaltens bzw. der Erwartungen von zumindest zwei Personen, die sich in einer bestimmten Situation befinden (Zwischenberger 2013:53). Denn um Qualität auf sachliche Weise

² Zu Rollenerwartungen im Bereich Konferenzdolmetschen siehe die Arbeit von Zwischenberger (2013).

beurteilen zu können, muss zunächst festgestellt werden, welche Bedeutungen den Dolmetscher_innen in einem bestimmten kulturellen Raum zukommen (Mack 2002:113).

In der Translationswissenschaft sind nur wenige Arbeiten zu finden, die sich mit den verschiedenen Auffassungen bzw. Definitionen von Qualität beschäftigen (Grbić 2008:233; Zwischenberger 2013:22). So soll als nächstes ein genauerer Blick auf die von Grbić (2008) und Zwischenberger (2013) formulierten Qualitätskonstrukte erfolgen.

Grbić (2008) geht bei der Beschreibung von Qualität als soziales Konstrukt von der Arbeit von Harvey & Green (1993) aus und bezieht Aspekte der drei sozialen Systeme Forschung, Didaktik sowie Praxis und deren Perspektiven von unterschiedlichen Zeitpunkten mit ein. Dabei beschreibt sie die drei sozialen Konstrukte *Quality as exception* (Qualität als Ausnahme), *Quality as perfection* (Qualität als Perfektion) und *Quality as fitness for a given purpose* (Qualität als Eignung für einen bestimmten Zweck), die wiederum in weitere Konstrukte unterteilt werden. Dabei widersprechen sich einige von ihnen, während andere koexistieren können oder sich gar ergänzen (Grbić 2008).

Nach Grbić (2008:241) basiert *Quality as exception* auf der Sichtweise, dass Qualität etwas Absolutes und Selbstverständliches ist, das eindeutig existiert und das man besitzen kann. Hierzu zählt beispielsweise die klassische Sicht (*The traditional notion*), der zufolge Qualität speziell und exklusiv ist. Sie wird instinktiv erkannt, braucht somit weder eine Definition oder Beschreibung noch Maßstäbe oder Kontrollmaßnahmen. Eine kleine Elitegruppe, umgeben von einem „figurative nimbus of mythical mastery“ (Grbić 2008:241) kontrolliert diese Qualität und genießt einen hohen sozialen Status. Dieses Konzept von Qualität ist in vielen frühen Beschreibungen über das Dolmetschen zu finden. Die Beschreibungen von Jean Herbert (1978) zählen hier beispielsweise dazu (Grbić 2008:241f.). So heißt es in einer seiner Beschreibungen: „The Secretary General was so impressed that he provided us with a very special rest-room equipped not only with comfortable armchairs, but also with four beds on which to relax.“ (Herbert 1978:7) Weiters fällt hier auch Qualität als exzellente Leistung durch das Übertreffen hoher Standards (*Excellence 1 or Exceeding high standards*) darunter. Auch dieses Konstrukt ist sehr elitär, da es sehr schwierig ist, über die hoch angesetzten Standards hinauszukommen, um zur Qualität zu gelangen. Hier fallen zum Beispiel die Sichtweisen von Eliteausbildungsstätten hinein, die nur die begabtesten Kandidat_innen aufnehmen. Als drittes gehört auch *Compliance with standards* zur Qualität als Ausnahme. Qualität ist gegeben, wenn klar definierte, als objektiv und statisch geltende, hohe Standards erfüllt werden, was jedoch schwer zu erreichen ist. Als Beispiel gilt hier die AIIC. Zwar steht es jeder Person prinzipiell frei, Teil des Systems zu werden, dessen Mitgliedschaft Prestige verspricht, doch ist es in der

Realität schwierig, die hohen Anforderungen zu erfüllen. So werden auch die Aufnahmevoraussetzungen der AIIC³ oftmals kritisiert (Grbić 2008:242ff.).

Quality as perfection hingegen folgt der Auffassung, dass Qualität durch persönliches Engagement und Anstrengung erreicht werden kann. Diese Sicht lässt sich wiederum in zwei konkretere Definitionen aufteilen. Fehlerfreie Exzellenz (*Excellence 2 or Zero defects*) will sicherstellen, dass alles korrekt und perfekt ist. Zwar entspricht Qualität gewissen Spezifikationen, sie kann jedoch nicht mit spezifischen Kriterien gemessen werden. Die Mitglieder des jeweiligen Systems sind für ihre Dolmetschungen verantwortlich und müssen sicherstellen, dass keine Fehler unterlaufen. Dieses Konstrukt konzentriert sich auf die Äquivalenz von AT und ZT und geht davon aus, dass Fehler vor allem auf der Mikroebene vorkommen. Diese Sichtweise legt den Fokus also auf das Dolmetschprodukt und ist im didaktischen Bereich zu finden, für den bestimmte Kriterien⁴ definiert werden. Das Konzept von Qualitätskultur (*Quality culture*) sieht Qualität als kollektive Aufgabe bzw. Mission aller Akteur_innen im System. Qualität soll demokratisiert werden, wobei jede Person zu jeder Zeit mitverantwortlich ist. Dabei wird Qualität relativiert, da es keine klar definierten Kriterien gibt, anhand derer man sie messen könnte, sondern lediglich eine Kultur, die nach Qualität strebt. Dies scheint den Bedürfnissen der heutigen Gesellschaft jedoch nicht gerecht zu werden (Grbić 2008:244-247).

In der letzten Gruppe der *Quality as fitness for purpose* führt Grbić (2008) drei Sichtweisen an. Diese gehen allesamt von der Frage aus, ob ein Produkt bzw. ein Service seinen Zweck erfüllen kann und ob die Kund_innen zufrieden gestellt werden können. Ist dies der Fall, spricht man von Qualität (Grbić 2008:247). Diese ist somit „neither elitist, nor exclusive, nor difficult to attain. Every product and every service is potentially useful and has the potential to satisfy the customer.” (Grbić 2008:247) Ausgehend davon, wie diese Nützlichkeit definiert wird, und wer sie bestimmt, können also drei Kategorien unterschieden werden. Die erste setzt Qualität mit der Zufriedenstellung der Bedürfnisse von Kund_innen gleich (*Usefulness for consumers 1 or The satisfaction of customer needs*). Dies setzt voraus, diese Bedürfnisse zu kennen und deren Erfüllung zu messen (Grbić 2008:247f.). Doch schon Shlesinger (1997) stellte die Frage: „Do our clients know what’s good for them?” (Shlesinger 1997:126) Wie

³ Eine Bewerbung ist zweimal in Jahr möglich, wobei einige Kriterien erfüllt werden müssen, darunter das Nachweisen von 150 Konferenztagen und drei bereits vollwertigen AIIC-Mitgliedern, mit mindestens fünfjähriger Mitgliedschaft, die die antragstellende Person als Sponsor_innen unterstützen, wobei zwei von ihnen ihren Berufswohnsitz in der gleichen Region haben müssen (AIIC 2019).

⁴ Die Bewertung in der Didaktik wird in Kapitel 2.3.3. behandelt, zu den Beurteilungskriterien im didaktischen Bereich siehe Kapitel 2.3.3.2.

bereits angemerkt, interessierte sich die Dolmetschwissenschaft seit den 1980er Jahren für die Perspektive der Kund_innen. Dennoch stellte sich bald heraus, dass deren Erwartungen sehr heterogen sind. Auf der einen Seite ist eine Orientierung an den Kund_innen wichtig, doch es bringt auch die Gefahr mit sich, ihnen ein Expert_innen-Wissen zuzuschreiben (Grbić 2008:248).

In der zweiten Kategorie des Konstruktes *Quality as fitness for purpose* rücken Marktmechanismen und Konkurrenz, sowie Angebot und Nachfrage in den Vordergrund. Qualität wird im Sinne eines Preis-Leistungs-Verhältnisses aufgefasst (*Usefulness for consumers 2 or Value for money*). So wird in Relation des Preises eine adäquate Dolmetschung erwartet. In der Realität fehlt oftmals das Verständnis für die geforderten Preise einer solchen Leistung, die nicht selten mit Sprachenkenntnis gleichgesetzt wird. Konkurrenz reguliert den Markt, kann aber auch zur Entstehung einer Preis-Dumping-Spirale führen (Grbić 2008:249f.). In der letzten Kategorie verlegt sich der Fokus auf die Rolle der Dienstleister_innen, die Qualität zu ihrer Mission erklären (*Usefulness determined by the service provider or Quality as mission*). Die Zufriedenstellung der Kund_innen ist immer noch das Ziel, doch sind es hier hauptsächlich die Dienstleister_innen, die Qualität definieren (Grbić 2008:250). Ein „high-quality service provider is one who states his/her mission on the basis of clearly defined standards and is effective in attaining the self-declared goals. Within this construct, fitness for purpose becomes fitness for the purpose of the market.“ (Grbić 2008:250) Hierbei spielen klare Kontrollmaßnahmen wie nationale und internationale Normen, beispielsweise in Form von Zertifikaten, eine zentrale Rolle (Grbić 2008:250).

Zwischenberger (2013) wiederum fasst das soziale Konstrukt der Qualität in zwei Metakonstrukten zusammen, die sie als „Qualitätssicherung und Qualitätsmanagement“ und „Qualitätsbeurteilung, Qualitätsbewertung und Qualitätsmessung“ bezeichnet. Zwischen diesen Metakonstrukten ist ein weiteres Konstrukt zu finden, nach dem sich Qualität aus einem holistischen Ansatz und Kooperation ergibt (Zwischenberger 2013:22,45).

Qualitätssicherung aus unternehmerischer Sicht, und mittlerweile als Unterkategorie zu Qualitätsmanagement definiert, hat zum Ziel, Vertrauen in den Bereich des Qualitätsmanagements zu schaffen (Zwischenberger 2013:22). So definiert die Internationale Organisation für Normierung (ISO) Qualitätssicherung als „part of *quality management* [...] focused on providing confidence that *quality requirements* [...] will be fulfilled“ (ISO 2015). Unter dem ersten Metakonstrukt „Qualitätssicherung und Qualitätsmanagement“ fasst Zwischenberger (2013) demnach solche Konstrukte zusammen, die sich auf die Voraussetzungen zur Qualitätssicherung bzw. Vertrauensschöpfung in Hinblick auf

Dolmetschqualität beziehen. Das erste Subkonstrukt „Qualität durch Qualifikation/Eignung“ bezieht sich auf die Anfangszeit des Konferenzdolmetschens, in der den Dolmetscher_innen mit Bewunderung begegnet wurde. Diese stellten eine kleine, elitäre Gruppe dar, die Dolmetschqualität garantierte und verkörperte. So bedurfte es auch keiner genaueren Definition von Qualität. Die Pionier_innen beschreiben die erforderlichen Eigenschaften von Dolmetscher_innen in ihren frühen Werken aus ihrer Perspektive (Zwischenberger 2013:22-25). Zweitens sieht das Konstrukt „Qualität durch Sinnerfassung“ die korrekte Erfassung und Weitergabe des von den Redner_innen gemeinten Sinns als zentralen Aspekt an. Dabei wird auf die Deverbalisierungstheorie von Seleskovitch (1976) Bezug genommen, die die drei Systeme Dolmetschwissenschaft, Didaktik und Praxis in ihren Anfängen besonders prägte. Auch wenn die Theorie mittlerweile auch kritisiert wird, ist die Idee von einem Sinn, der ohne Probleme erfasst werden kann, noch heute in der Praxis sowie Didaktik des Dolmetschens präsent. So wird auch durch die AIIC das Erschließen des von den Sprecher_innen gemeinten Sinns als primäre Aufgabe von Konferenzdolmetscher_innen angesehen (Zwischenberger 2013:25ff.). Das dritte Subkonstrukt zur Qualitätssicherung wurde von Zwischenberger (2013) mit „Qualität als Erfüllung hoher professioneller Voraussetzungen“ benannt. Dieses Konstrukt wird vor allem von der AIIC propagiert. Zwar steht es jeder Person offen, Mitglied dieses Verbandes zu werden, doch stellen die hohen Anforderungen eine große Hürde dar. Als Mitglied genießt man jedoch Prestige und sozialen Status, während Außenstehende die Aufnahmeveraussetzungen oftmals als diskriminierend ansehen (Zwischenberger 2013:27ff.). Im Gegensatz dazu ist es Ziel des Subkonstrukts „Qualität als Normerfüllung“, Mindestanforderungen zu definieren, um die Qualität des Dolmetschens durch ein zufriedenstellendes, akzeptables Mindestmaß zu sichern. Dazu werden in verschiedenen Ländern bzw. für die verschiedenen Märkte unterschiedliche Normen formuliert, die sich in ihren Spezifikationen und den Bereichen, die sie abdecken, unterscheiden. Für Österreich sind für Dolmetschnormen die D 1202 und D 1203 zu nennen. Solche Normen bzw. Dokumente beinhalten jedoch keine konkreten Maßnahmen, die für eine tatsächliche Überprüfung solcher Mindestanforderungen bzw. der Dolmetschqualität herangezogen werden könnten. Solche Maßnahmen sind im folgenden Metakonstrukt enthalten (Zwischenberger 2013:29ff.).

Das Metakonstrukt „Qualitätsbeurteilung, Qualitätsbewertung und Qualitätsmessung“ umfasst nach Zwischenberger (2013) vier Konstrukte. Dabei werden die Begriffe Bewertung und Beurteilung synonym verwendet und beziehen sich auf Leistungen von professionellen Dolmetscher_innen und Student_innen während Messung für die Evaluation in experimentellen Situationen verwendet wird (Zwischenberger 2013:32). Wie bereits in Kapitel 2.1. erläutert,

wurden die ersten Studien zum Simultandolmetschen in experimentellen Untersuchungen aus der Perspektive anderer Disziplinen durchgeführt, wobei der Fokus auf der Mikroebene und der Messung linguistischer Übereinstimmung lag. Diese Sichtweise fasst Zwischenberger (2013) als das Konstrukt „Qualität als Ergebnis einer Messung der sprachlichen Korrespondenz“ auf und verweist auf experimentelle Untersuchungen ab den 1960er Jahren. Auch wenn die Qualitätsauffassung dieses Konstruktes heute stark kritisiert wird, hatte sie dennoch großen Einfluss (Zwischenberger 2013:32ff.). Das zweite Konstrukt „Qualität als Mangelfreiheit“ ist im didaktischen Bereich des Simultandolmetschens verortet und kann wiederum in zwei Ansätze unterteilt werden. Der erste stellt Qualität durch eine Analyse basierend auf zuvor definierten Fehlern fest, die nicht gemacht werden dürfen, während der zweite Ansatz von Seiten der Studierenden erwartet, gewisse Parameter zu erfüllen. In dieses Konstrukt fallen somit unterschiedliche Methoden und Schemata für Evaluationen in der Didaktik⁵, wie sie beispielsweise von Riccardi (2003) oder Kutz (2005) entwickelt wurden (Zwischenberger 2013:34-37). Das dritte Konstrukt in dieser Gruppe stellt die „Qualität als Ergebnis von Nutzerbewertungen“ dar. An dieser Stelle ist besonders Kurz (2003) zu nennen, die diese Sichtweise anhand einer Formel darstellt (Zwischenberger 2013:37f.). In dieser wird kurz festgehalten: „Quality = Actual Service – Expected Service“ (Kurz 2003:17). Nach dieser Sichtweise richtet sich Qualität ausschließlich danach, die Erwartungen der Nutzer_innen⁶ zu erfüllen. Eine große Anzahl empirischer Arbeiten konstruierten dieses Konstrukt mit. Von der Studie Bühlers (1986) ausgehend wurde bald darauf von Kurz (1989) die erste Untersuchung durchgeführt, bei der Nutzer_innen befragt wurden. Diese führten zu einer Reihe weiterer Befragungen, um ihre Erwartungen zu eruieren, die sich als äußerst heterogen herausstellten. Doch auch Studien, in denen Nutzer_innen konkrete Dolmetschungen bewerten sollten, wurden durchgeführt. Es wurde zudem festgestellt, dass sich die Gruppe der Nutzer_innen und jene der Dolmetscher_innen in Hinblick auf ihre Erwartungen unterscheiden (Zwischenberger 2013:37-42). Wenn Qualität im Gegensatz zu diesem Konstrukt nach dem Erfüllen von Erwartungen praktizierender Dolmetscher_innen definiert wird, handelt es sich nach Zwischenberger (2013) um das Konstrukt „Qualität als Resultat von Expertenbewertungen“ (Zwischenberger 2013:42-45).

Zwischenberger (2013) formuliert ein weiteres Konstrukt, das eine holistische Sicht darstellt. Sie verortet dieses zwischen den beiden bereits beschriebenen Metakonstrukte, da es

⁵ An dieser Stelle sei nochmals darauf verwiesen, dass das Kapitel 2.3.3. sich mit der Evaluation im didaktischen Bereich auseinandersetzt.

⁶ Auf Nutzer_innen bzw. Rezipient_innen wird in Kapitel 2.3.2. bzw. 2.3.2.2. näher eingegangen.

hierbei sowohl darum geht, Qualität zu sichern, als auch Beurteilungen durchzuführen. In diesem Konstrukt sind die Auffassung von Konferenzen als Hypertexte nach Pöchhacker (1994) und das Anforderungsmodell von Kalina (2004), die bereits in Kapitel 2.1. aufgegriffenen wurden, zu verorten. Bei diesem Konstrukt kommt dem Kontext⁷, der Situation und den Arbeitsbedingungen große Wichtigkeit zu (Zwischenberger 2013:45-48).

Nach diesem Überblick über die Qualitätskonstrukte, wie sie in der Translationswissenschaft beschrieben werden, wird klar, dass einige Konstrukte von Grbić (2008) und Zwischenberger (2013) ähnlich, andere unterschiedliche eingeteilt werden. An dieser Stelle sei daher noch auf einen Punkt hingewiesen. Grbić (2008) hebt hervor, dass es sich bei ihrer Einteilung um keine stabile Typologie handelt und kein Anspruch auf eine erschöpfende Darstellung erhoben wird (Grbić 2008:251). Es handelt sich eher um „an attempt to explore a complex subject from different perspectives“ (Grbić 2008:251). Auch Zwischenberger (2013) merkt an, dass ihre Art, das soziale Konstrukt der Qualität in der Translationswissenschaft darzustellen, wiederum selbst als Konstruktion angesehen werden muss und eine zeit- und kulturgebundene Perspektive darstellt (Zwischenberger 2013:20). Dem folgend soll an dieser Stelle auch festgehalten werden, dass auch diese Arbeit in dem Bewusstsein entstanden ist, dass die vorliegende Auswahl und Aufbereitung der präsentierten Erkenntnisse und Theorien im großen Hauptkapitel „Qualität“ nur eine von vielen Möglichkeiten sind, um das Thema der Dolmetschqualität aufzuarbeiten.

Die bisherigen Erläuterungen führen nochmals deutlich vor Augen, dass es bis heute nicht gelungen ist, eine allgemeine Definition von Qualität aufzustellen. Die beschriebenen Qualitätskonstrukte sind „geschichts- und kulturgebunden und fungieren als semantische Klammern, die die Vielzahl an Perspektiven, Interessen, Intentionen und Konzepten der VertreterInnen unterschiedlicher sozialer Systeme zum Ausdruck bringen“ (Zwischenberger 2013:51). Diese Verschiedenheit der Vorstellungen schlägt sich in weiterer Folge in den verschiedenen Erwartungen nieder, die an die Rolle bzw. Rollen der Dolmetscher_innen gestellt werden (Zwischenberger 2013:51).

Schließlich sind auch die Auffassungen und Perspektiven auf das Dolmetschen selbst entscheidend und beeinflussen in weiterer Folge die Evaluation von Qualität (Mack 2002:113f.). Denn

[f]asst man Dolmetschen als Austausch von Textmaterial zweier Sprachen auf, so ist seine Qualität eine Funktion der Fähigkeit zur Herstellung von Äquivalenz hinsichtlich Inhalt, Form

⁷ Auf die Bedeutung des Kontexts wird in Kapitel 2.3.2. bzw. 2.3.2.1. näher eingegangen.

und allenfalls Darbietungsart; betrachtet man es hingegen als Sonderform interlingualen kommunikativen Handelns in einem komplexen sozialen Beziehungsnetz, so resultiert seine Qualität vielmehr aus der Optimierung pragmatisch bedingter Sprechhandlungen. (Mack 2002:114)

Dabei existieren zahllose Definitionen vom Dolmetschen. Doch nur unter Berücksichtigung des Wesens des Dolmetschens selbst und unter Einbeziehung seiner Funktionen und der erwarteten Ziele kann Dolmetschqualität evaluiert werden (Viezzi 2007:172).

Es wird also deutlich, dass es sich bei Qualität um ein Konzept „in constant evolution“ (Alonso Bacigalupe 2013:10) handelt. Mit der Zeit änderten sich die Prioritäten und neue Prinzipien und Konzepte kamen hinzu (Alonso Bacigalupe 2013:10). Somit ist Qualität „a social construct generated by a particular time and its people, the context and culture in which it is evaluated“ (Pradas Macías & Zwischenberger 2022:248). Das nachfolgende Kapitel ist der Qualitätsbeurteilung gewidmet und geht nun genauer auf die Aspekte ein, die hierbei eine Rolle spielen.

2.3. Evaluation von Qualität

Bis heute erweist sich die Evaluation von Verdolmetschungen als schwierig (Behr 2013a:15) und stellt ein kontroverses Unterfangen dar (Riccardi 2002:116). Wie bereits erläutert, stellt Qualität keine neutrale Eigenschaft dar und ist einer Sache nicht inhärent. Ihre Bewertung hängt sowohl von der Situation als auch den Personen ab, die evaluieren (Grbić 2008:234). Auch die herangezogenen Maßstäbe für Qualität sind der zu beurteilenden Interaktion nicht immanent, sondern werden von den unterschiedlichen Akteur_innen und Systemen definiert bzw. konstruiert „in order to satisfy the everyday needs of the respective system and to control its participants' behaviour in order to ensure that the system continues to function and that quality is achieved“ (Grbić 2008:240f.). So variiert Qualität „je nachdem von wem, wann, wo, wie, in welcher Situation und unter welchen Bedingungen sie beurteilt [...] wird“ (Zwischenberger 2013:17).

Mittlerweile herrscht also Konsens darüber, dass Qualitätsbewertung höchst variabel ist, und von einer Vielzahl an Perspektiven sowie einer unendlichen Anzahl an kontextuellen Parametern abhängig ist (Grbić 2008:237). Kutz (2005) spricht von einer Bewertungsvariabilität, die über die spezifischen Perspektiven hinausgeht und u.a. auch die unterschiedliche Schwerpunktsetzung bei der Bewertung berücksichtigt. Er sieht das Bewerten

als Handlungstyp, bei dem die große Bandbreite an einzelnen Kriterien berücksichtigt wird (Kutz 2005:15), denn

[d]ie Palette reicht von der Situationsangemessenheit des Äußeren (einschließlich der Frisur und der Bekleidung) über die Prosodie, den Höreindruck der Stimme des Dolmetschers, über die Funktionskonstanz der Inhaltswiedergabe bis hin zu feinen Schattierungen der sprachlichen Realisierung der Dolmetschleistung und den durch den Dolmetscher zu meisternden Problemlösungsfällen usw. (Kutz 2005:15)

Moser-Mercer (1996:45) unterscheidet in diesem Kontext zwischen den drei Begriffen *evaluation*, *measurement* und *assessment*, wobei deren Verwendung davon abhängt, in welchem Kontext ein Blick auf Qualität geworfen wird. Sie erläutert hierzu:

[I]f we are concerned with the quality of interpreting in a natural setting, we probably want to evaluate it. If, however, our goal is to analyze the product of interpreting in a laboratory setting, we would most likely proceed to measurement. In working with student interpreters, [...] assessment would be the proper approach, until, at least, the student's final course exams, when we are more likely going to combine both evaluation and assessment. (Moser-Mercer 1996:45)

Eine vergleichbare Unterscheidung für Dolmetschungen von professionellen Dolmetscher_innen und Student_innen kann im Deutschen anhand der verfügbaren Begrifflichkeiten nicht getroffen werden (Zwischenberger 2013:32). In dieser Arbeit werden daher die Begriffe Evaluation, Evaluierung, Bewertung sowie Beurteilung synonym verwendet, ohne einen expliziten Unterschied zu machen.

Im vorliegenden Kapitel soll zunächst ein allgemeiner Blick auf die Qualitätsparameter zur Evaluation von Dolmetschleistungen geworfen werden, bevor näher auf Kontext und Bewertungsperspektive eingegangen wird. Da das Forschungsinteresse der vorliegenden Arbeit im didaktischen Kontext zu verorten ist, wird diesem Bereich ein eigener Teil gewidmet, um die Bewertungsarten und die verwendeten Kriterien in der Dolmetschdidaktik näher zu beleuchten.

2.3.1. Kriterien der Qualitätsbewertung

Betrachtet man Qualität als eine Eigenschaft, die nicht beobachtbar ist, sondern aus einer Bewertung resultiert, wird deutlich, dass auch die herangezogenen Evaluationskriterien aus den impliziten oder expliziten Entscheidungen der Personen resultieren, die diese Qualität

sicherstellen oder feststellen wollen (Heid 2000:41). Dies ist auch dort der Fall, „wo bereits geltende, tradierte oder konventionelle und insofern vorgefundene Qualitätskriterien, die allerdings ihrerseits aus Entscheidungen resultieren, in einer vielleicht unreflektierten, aber nicht dispensierbaren Entscheidung ‚lediglich‘ übernommen werden“ (Heid 2000:41). So konnte bis heute keine allgemeine Rangfolge von Kriterien mit einer allgemein gültigen Gewichtung aufgestellt werden (Kalina 2004:3).

In den ersten Untersuchungen zur Qualität wurden Kriterien wie das *décalage* und Sprechpausen untersucht. Man nahm an, dass nur dadurch Dolmetschen überhaupt erst möglich war. In den frühen Arbeiten zum Vergleich von AT und ZT wurden, wie bereits angedeutet, lediglich Merkmale auf der Mikroebene berücksichtigt. Man untersuchte Fehler, während die Kommunikationssituation außer Acht gelassen wurde. Auch Aspekte auf der Makroebene, wie beispielsweise Kohärenz, wurden nicht miteinbezogen (Kalina 2004:3). All „die Texteigenschaften, die Texte überhaupt dolmetschbar machen, [blieben] noch unberücksichtigt“ (Kalina 2004:3). Den Fokus rein auf die Wortebene zu setzen ist jedoch kontraproduktiv, da so außer Acht gelassen wird, ob die Botschaft richtig übermittelt wurde (Sunnari 2003:235).

In den Anfängen der Dolmetschwissenschaft war die Qualitätsbewertung subjektiv und sehr intuitiv. Durch die vorherrschende Deverbalisierungstheorie wurde Qualität am Sinn festgemacht, was methodisch schwer zu überprüfen war. So spielten Empfindungen und Erfahrung in die Definition und in weiterer Folge in die Beurteilung der Qualität hinein (Kalina 2004:2). Für eine längere Zeit war Qualität „etwas implizites, etwas was man sofort erkennt, aber schwer mit Worten fassen kann; etwas, das sich einer genauen Erklärung entzieht“ (Riccardi 2005:99). Daher wurden auch sehr generelle Parameter, wie beispielsweise Texttreue, Klarheit, Äquivalenz oder Kohäsion verwendet, um Qualität zu definieren (Riccardi 2005:99). So galten ebenso eine universitäre Ausbildung bzw. die Mitgliedschaft in einem Verband wie beispielsweise der AIIC als Indikatoren für Dolmetschqualität. Die Abwendung von dieser Sichtweise brachte jedoch das Bedürfnis mit sich, explizite Qualitätskriterien zu formulieren (Pöchhacker 2004:153).

Durch die Studie von Bühler (1986) wurden Qualitätskriterien für die Evaluation von Verdolmetschungen zum Kernthema der Qualitätsforschung im Dolmetschen (Pöchhacker 2012:21). Der Fragebogen in Bühlers (1986) Untersuchung umfasste die Parameter Stimme, Flüssigkeit, Akzent, logische Kohäsion, angemessener Stil, korrekte Terminologie, Sinnübereinstimmung, Vollständigkeit der Informationen, korrekte Grammatik, Vorbereitung, Ausdauer, Haltung, angenehme Erscheinung, Zuverlässigkeit, Teamwork und positives

Feedback (Bühler 1986). Zunächst hat man versucht, Qualitätskriterien durch die Befragungen von Rezipient_innen aufzustellen. In den verschiedenen Arbeiten zu diesem Thema variiert die Anzahl der herausgearbeiteten Kriterien. Diese werden zudem teils detaillierter aufgelistet, teils auch allgemeiner gehalten. Allgemein kristallisierte sich dabei die Unterscheidung bzw. Unterteilung in kommunikativ-sprachliche und inhaltsbezogene Parameter heraus (Riccardi 2005:104-107). Einen Einblick in die Bewertungskriterien, die in der Literatur zu finden sind, bietet beispielsweise Fünfer (2013) in einem tabellarischen Überblick. Die angeführten Studien umfassten jeweils eine unterschiedliche Anzahl von verwendeten Kriterien (Fünfer 2013:40f.).

Die Tatsache, dass Definitionsversuche zu Qualität vage geblieben sind, führte zu einer Unzufriedenheit, die wiederum zu einer Auseinandersetzung mit spezifischen Einzelaspekten führte. Diese Untersuchungen führten so zu einer Zunahme von Qualitätskriterien (Behr 2013a:54). Dies wird beispielsweise durch den Sammelband von Collados Aís et al. (2011) deutlich, der einen interdisziplinären Blick auf zehn Parameter der Qualität für den Bereich des Simultandolmetschens bietet. Dieser beinhaltet die Parameter Eindrucksbildung, Stimme, Intonation, Flüssigkeit, Diktion, Akzent, logische Kohäsion, Stil, Terminologie und korrekte und vollständige Sinnübertragung (Collados Aís et al. 2011). Diese Zunahme an Kriterien ist jedoch widersprüchlich zu einer praktikablen Nutzung bei der Evaluation, auch wenn sie alle notwendig sind, um das Phänomen Dolmetschen und seine Qualität vollständig zu erfassen (Behr 2013a:54). Behr (2013b) merkt an dieser Stelle an, dass bei der Diskussion über die Frage, was die wirklich relevanten Qualitätskriterien seien, übersehen wird, dass es bereits einen Konsens gibt. In der Praxis liegt die Anzahl der verwendeten Kriterien zwischen drei und fünf. Unter Berücksichtigung des Settings werden prinzipiell die Kriterien inhaltliche Vollständigkeit, richtige und genaue Aussage und angemessene Präsentation angewendet (Behr 2013b:87). Oder „[n]och knapper gefasst könnte man sagen, eine gute Verdolmetschung wird dem situativen Kontext entsprechend funktional erbracht“ (Behr 2013b:87). Unter Einbeziehung von mehr Evaluationskriterien ist eine Durchführung nicht mehr realistisch. Konkreter ausformulierte Kriterien basieren auf jenen von Bühler (1986) bzw. jenen von Riccardi (2005) (Behr 2013b:87).

Riccardi (2002) erläutert Makrokriterien, die beim Bewerten von Dolmetschungen zum Einsatz kommen. Diese setzen sich wiederum aus Mikrokriterien zusammen, die jede Dolmetschleistung beeinflussen. Letztere sind es, die es im Studium zu erlernen gilt, weshalb sie auch für die Beurteilung von studentischen Leistungen herangezogen werden, während professionelle Leistungen anhand von Makrokriterien zu evaluieren sind. Hierbei bezieht sie sich auf die von Viezzi (1996) formulierten Kriterien bzw. Qualitätsziele in Hinblick auf

Dolmetschleistungen. Diese umfassen die Ziele *equivalence*, *usability*, *accuracy* und *appropriateness* (Riccardi 2002:117f.). Dabei sieht Viezzi (2007) das Dolmetschen als „a function of the attainment of goals“ (Viezzi 2007:172) und definiert es als eine Dienstleistung, die als textproduzierende und sprachen- und kulturübergreifende Tätigkeit angesehen werden kann (Viezzi 2007:172f.). Diese Makrokriterien stellen jedoch nur ein Beispiel dar. Auch andere Dolmetschwissenschaftler_innen haben allgemeine Kriterien für die Qualität von Dolmetschungen beschrieben. Diese ermöglichen auf der einen Seite einen Überblick über die Texteigenschaften einer Dolmetschung, können diese auf der anderen Seite jedoch nicht differenziert erfassen. Mikrokriterien bringen hingegen das Problem mit sich, dass das Gesamtbild der Verdolmetschung durch stark gegliederte Bewertungssysteme erschwert wird (Riccardi 2007:191-198).

Da laut Behr (2013b) die gängigen, in der Praxis verwendeten Kriterien auf den Ausführungen von Riccardi (2005) beruhen (Behr 2013b:87), sollen diese näher ausgeführt werden. Riccardi (2005) stellt vier Makrokriterien vor, die im Laufe ihrer Arbeiten und langjähriger Tätigkeit entwickelt wurden. Dabei spricht sie von der Textsorte des simultandolmetschten Texts, auf der diese Kriterien basieren. Sie sind jeweils in fünf weitere Klassen unterteilt. Jeder Makrobereich kann durch bestimmte Akteur_innen überprüft werden. Die ersten zwei Makrokriterien „Darbietungsart“ und „Sprachebene“ beziehen sich auf Aspekte der Textoberfläche und können sowohl von den Dolmetscher_innen und Dolmetschforscher_innen als auch von den Rezipient_innen überprüft werden. Die Kriterien des dritten Bereichs „Inhaltsentsprechung“, der insbesondere die Äquivalenzbeziehungen zwischen dem AT und dem ZT umfasst, sowie jene des Bereichs „Dolmetschstrategien“ können jedoch nur von den Simultandolmetscher_innen und den Forscher_innen angewendet werden. Die relativen Gewichtungen der beschriebenen Kriterien hängen von der Kommunikationssituation ab, sowie davon, wie oft die jeweiligen Aspekte vorkommen. So erschwert beispielsweise ein zu häufiges Vorkommen der Aspekte aus den Makrobereichen zur Textoberfläche die Rezeption der Verdolmetschung (Riccardi 2005:107-110). Tabelle 1 zeigt die vier Makrobereiche und ihre einzelnen Kriterien nach Riccardi (2005).

Tabelle 1: Qualitätskriterien nach Riccardi (2005:108ff.)

Makrobereich Darbietungsart	Makrobereich Inhaltsentsprechung
1. Akzent und Phonetik 2. Output 3. Prosodie 4. Verzögerungsphänomene 5. Pausen	1. Ersatzinformation 2. Auslassungen 3. Zusatzinformation 4. Logische Verbindungen 5. Register
Makrobereich Sprachebene	Makrobereich Dolmetschstrategien
1. Allgemeinwortschatz 2. Terminologie 3. Interferenz 4. Internationalismen 5. Morphosyntax und Syntax	1. Reformulierung 2. Antizipation 3. Décalage 4. Technik 5. Gesamtleistung

Im Makrobereich „Darbietungsart“ bezieht sich die Unterklasse „Akzent und Phonetik“ bei Riccardi (2005) auf die Abweichungen von der als Norm geltenden Aussprache und auf das Vertauschen von Lauten und Phonemen. Die Sprechgeschwindigkeit und das Sprechtempo fallen unter „Output“, während „Prosodie“ den Wort- sowie Satzakkzent und Abweichungen von der Intonation umfasst. Der vierte Punkt „Verzögerungsphänomene“ kann in zwei weitere Unterklassen geteilt werden. Die erste umfasst Aspekte wie beispielsweise hörbares Atmen oder Häsitationslaute, die zweite bezieht sich auf Neuplanungen und Abbrüche. Schließlich fallen unter das Kriterium „Pausen“ solche ungefüllten Pausen ab einer Länge von 3 Sekunden, sofern sie nicht im Ausgangstext vorkommen (Riccardi 2005:108).

In den Bereich „Sprachebene“ fallen lexikalische, terminologische und morphosyntaktische Aspekte hinein. Der „Allgemeinwortschatz“ und die „Terminologie“ beziehen sich dabei sowohl auf Kollokationen als auch einzelne Wörter. Durch das Kriterium der „Interferenz“ wird die Verdolmetschung auf eine zu hohe Anzahl von Lehnübersetzungen aus der Sprache des Ausgangstextes überprüft, die kommunikationsstörend sein können. Unter „Internationalismen“ werden Transliterationen erhoben. Schließlich wird mit dem Kriterium „Morphosyntax und Syntax“ das Abweichen bei der Kohäsion und Konkordanz erfasst (Riccardi 2005:108).

Der dritte Makrobereich überprüft, inwiefern die Informationseinheiten aus dem Ausgangstext und deren Wechselbeziehungen in der Dolmetschung zu finden sind. Unter diesen Bereich fallen zunächst die Texterscheinungen „Ersatzinformation“, „Auslassungen“ und „Zusatzinformation“. Diese können jeweils in weitere Kategorien unterteilt werden (Riccardi 2005:109). Prinzipiell sind sie nämlich „weder positiv noch negativ, da sie erst im Zusammenhang mit der Kommunikationssituation und dem Kontext ihr relatives Gewicht erhalten“ (Riccardi 2005:109), und somit sowohl kohärenzfördernd als auch kohärenzschädlich sein können. Zur Überprüfung, ob Struktur und Logik der Originalrede beibehalten wurden, dient die Klasse „Logische Verbindungen“. Schließlich kann mittels des letzten Kriteriums das „Register“ untersucht bzw. mit jenem des Ausgangstextes verglichen werden, um zu beurteilen, ob es möglicherweise zu einer Reduzierung oder Verstärkung der kommunikativen Intention der Sprecher_innen gekommen ist (Riccardi 2005:109).

Unter „Dolmetschstrategien“ fallen zunächst vier Kriterien, mittels derer die Dolmetschkompetenz erfasst werden kann. „Reformulierung“ bezieht sich dabei auf die Fähigkeit, losgelöst von der Ausgangssprache Lösungen in der Zielsprache zu finden, die angemessen und äquivalent sind. Eine weitere Strategie stellt die „Antizipation“ dar, die durch Hypothesen und Schlussfolgerungen möglich ist. Mit dem „Décalage“ wird der Abstand zwischen dem Sprechen der Redner_innen und der Dolmetscher_innen erfasst. Dies kann beispielsweise bei Debatten oder Podiumsdiskussionen oder im Falle von Relaisdolmetschungen negative Auswirkungen haben. Die Unterklasse „Technik“ bezieht sich auf die Technik des Simultandolmetschens. Diese umfasst Aspekte wie Selfmonitoring oder geteilte Aufmerksamkeit und das richtige Einsetzen der Technik. Aber auch die Fähigkeit, in schwierigen Situationen strategisch zu handeln, fällt hier hinein. Im letzten Punkt dieses Makrobereiches verortet Riccardi (2005) die „Gesamtleistung“. Diese überprüft die Erreichung des Kommunikationszweckes, indem der Gesamteindruck auf die Zuhörer_innen direkt nach dem Ende der Dolmetschung erfasst wird (Riccardi 2005:110). Diese letzte Klasse „umfasst alle anderen Klassen, indem sie einzelne Aspekte subsumiert und erhält dadurch einen eigenen Wert“ (Riccardi 2005:110).

Dennoch kann zusammenfassend gesagt werden, dass es schwierig ist, Kriterien für die Evaluation zu finden, die auf alle Settings übertragbar wären (Behr 2013a:28). Denn eine Dolmetschung ist stets „linked to the temporal and spatial circumstances in which it takes place“ (Riccardi 2002:116). Zusätzlich stellt sich bei einer Evaluation die Frage, aus welcher Perspektive bewertet werden soll (Pöchhacker 1994:123). Daher soll die Rolle des Kontextes und der Perspektive im Folgenden näher beleuchtet werden.

2.3.2. Kontext und Bewertungsperspektive

Dolmetschen findet nicht im Vakuum statt: Dolmetscher_innen agieren stets in einem komplexen sozialen Gefüge von Handlungen (Mack 2002:117). Riccardi (2005:103) spricht vom simultangedolmetschten Text als eigene Textsorte. Dieser „ist in der kommunikativen Situation völlig eingetaucht“ (Riccardi 2005:103). Die Besonderheit der Dolmetschleistung besteht darin, dass sie „naturgemäß dazu bestimmt [ist], einmalig, und zwar unmittelbar im Zuge ihrer Erbringung, verwendet zu werden“ (Mack 2002:111). Somit ist Dolmetschen flüchtig und immer in eine komplexe Kommunikationssituation eingebunden (Mack 2002:111). Diese wird von unterschiedlichen Faktoren beeinflusst, wie beispielsweise der Art des Treffens und dem Thema, das behandelt wird. Aber auch die Art und Weise, wie die Texte dargeboten werden, ob sie vorgelesen werden oder als spontane Rede gehalten werden, sowie ihr Grad an Fachlichkeit spielen eine Rolle. Schließlich stellen auch die beteiligten Personen wie die Redner_innen und Zuhörer_innen einen Teil der Situation dar (Riccardi 2005:101). Bei der Bewertung müssen daher auch stets die Umstände, unter denen die Simultandolmetschung erbracht wurde, Berücksichtigung finden (Riccardi 2005:104).

Eine Schwierigkeit besteht darin, dass die verschiedenen Akteur_innen, die einen Teilaspekt der Qualität darstellen, auch unterschiedliche Sichtweisen auf Qualität haben und diese unterschiedlich wahrnehmen (Garzone 2003:23). Denn „quality may mean very different things to different persons“ (Kalina 2005:771). So können auch die unterschiedlichen Qualitätskriterien aus den Blickwinkeln der verschiedenen Akteur_innen untersucht werden, was wiederum die Anzahl der zu berücksichtigenden Parameter erhöht (Garzone 2003:23). Abgesehen davon, dass die Perspektive auf Qualität vom Kontext und der jeweiligen Person abhängt, bringt ihr relativer Charakter auch mit sich, dass dieselben Personen den Begriff zu unterschiedlichen Zeitpunkten möglicherweise auch auf unterschiedliche Art und Weise verwenden (Harvey & Green 2000:17). Die Qualität liegt somit „im Auge des Betrachters (bzw. im Ohr des Zuhörers)“ (Pöchhacker 1994:123).

An dieser Stelle stellt sich die Frage, wer die Bewertung vornimmt. Bereits Bühler (1986) erläutert auf die Frage, wer bewerten sollte, dass es zumindest fünf mögliche Antworten gäbe. So können Dolmetschungen aus Sicht von Lehrenden, professionellen Organisationen, Kolleg_innen, Auftraggeber_innen oder Endnutzer_innen beurteilt werden (Bühler 1986:231). Aber auch die Textproduzent_innen, die die eigene Leistung einschätzen, oder beobachtende Dritte ohne direktes kommunikatives Interesse in Hinblick auf die Verdolmetschung, kommen als Beurteilende in Frage. Letztere können neben Kolleg_innen der Dolmetscher_innen auch

Vertreter_innen der Auftraggeber_innen oder Redner_innen sein, sowie Dolmetschwissenschaftler_innen. Während die Gruppe der Dolmetschwissenschaftler_innen am stärksten nach einer objektiven und intersubjektiven Beurteilung strebt, stellen die Zuhörer_innen die Perspektive der direkten subjektiven Beurteilung dar (Pöchhacker 1994:123f). Pöchhacker (1994:124) räumt den Auftraggeber_innen bzw. deren Vertreter_innen durchaus ein, sich ein Urteil über die Qualität bilden zu können, das beispielsweise auf einem reibungslosen Ablauf basiert (Pöchhacker 1994:124). Diese Gruppe hat kein „direktes Interesse am einzelnen Text [...], sondern an der Qualität der Handlungen und Handlungsprodukte in der Hypertextdimension“ (Pöchhacker 1994:124). Das Interesse an Qualität und dessen Bewertung aus Sicht der Konferenzveranstalter_innen ist somit funktionaler Natur. Im Mittelpunkt steht die Frage, ob der Veranstaltungsverlauf für die teilnehmenden Personen konfliktlos und nützlich war (Kutz 2005:14). Da aber jede Perspektive unausweichlich subjektiv ist, müssen die unterschiedlichen Urteile aus den verschiedenen Blickwinkeln miteinander in Beziehung gesetzt werden. Denn beispielsweise die Evaluation durch Kolleg_innen stellt trotz fachlicher Kompetenz auch eine im Prinzip subjektive Perspektive dar (Pöchhacker 1994:124). Zusätzlich spielen auch Aspekte wie Verständnis und Teamgeist auf der einen Seite, und Wettbewerbsmomente auf der anderen Seite eine Rolle in dieser Perspektive. Schließlich ist eine mögliche Bewertungsperspektive die der Dolmetscher_innen selbst. Sie ist subjektiver Natur und richtet sich nach dem individuellen Eindruck und dem Feedback durch das Publikum (Kutz 2005:15).

Im Folgenden soll nochmals differenzierter auf die Bedeutung des Kontexts eingegangen werden, bevor genauer auf die Rezipient_innen eingegangen wird, die einen zentralen Aspekt der vorliegenden Arbeit darstellen.

2.3.2.1. Die Bedeutung des situativen Kontextes

Wie bereits erläutert, wurde in den 1980er Jahren die Rolle des situativen Kontextes in der Dolmetschwissenschaft erkannt (Kalina 2004:4). Doch bereits Herbert (1952:82f.) weist auf die Bedeutung der Situation hin und meint,

[i]t is quite clear that in a diplomatic conference the greatest attention should be paid to all the nuances of words, while in a gathering of scholars, technical accuracy will have greater importance; in a literary and artistic gathering, elegance of speech; and in a political assembly, forcefulness of expression. (Herbert 1952:82f.)

Pöchhacker (1994) beschreibt sieben Konferenztypen bzw. Hypertext-Prototypen. Zunächst nennt er Versammlungen internationaler Organisationen, die sich sowohl auf zwischenstaatliche als auch nichtstaatliche bzw. private Organisationen beziehen. Beim zweiten Typ handelt es sich um Fachkonferenzen, gefolgt von Seminaren und Schulungen. Verhandlungen stellen den vierten Punkt dar und meinen genauer gesagt Arbeitstagungen von Gremien, beispielsweise in Form von Kommissionen oder Ausschüssen. Der Typ „Aktuelles Forum“ umfasst Diskussionsveranstaltungen, die sich meistens auf aktuelle politische Themen beziehen und international besetzt sind. Die zwei letzten Typen stellen Pressekonferenzen und Präsentationen, sowie Gastvorträge dar. Zusätzlich können diese Konferenztypen durch fünf Merkmale differenziert werden. Hierzu zählen das Ausmaß der Strukturierung, die Homogenität der Gruppe, die Informationsintensität, das genutzte visuelle Begleitmaterial und der Informationsfluss bzw. die Dynamik (Pöchhacker 1994: 51f.).

Kutz (2005) nennt unterschiedliche Situationstypen, in die ein bestimmtes Kommunikationsereignis je nach den Bestimmungen des Auftrags eingeteilt werden kann. Eine Dolmetschsituation kann formbetont und inhaltsbetont sein. Bei solchen Aufträgen spielen protokollarische und ästhetische Aspekte sowie Massenwirksamkeit eine zentrale Rolle. Hierin fallen zum Beispiel Empfänge, Lesungen von Schriftsteller_innen oder Kundgebungen. Bei inhaltsbetonten Situationen sollen vor allem die Informationen genau wiedergegeben werden. Die sprachliche Gestaltung soll in einer funktionalen Art und Weise geschehen. Verfügen die Zuhörer_innen über geringeres Sachwissen können zusätzliche Informationen zu bestimmten Begriffen oder Zusammenfassungen erforderlich sein. In diese zweite Gruppe fallen Veranstaltungen wie Konferenzen, Seminare oder Fachtagungen. Bei der dritten Gruppe handelt es sich um rein pragmatische Situationen. Diese Dolmetschsituationen zeichnen sich durch eine dramatische und zugespitzte Lage aus. Diese Kategorie umfasst das Dolmetschen bei der Polizei, sowie kommerzielles und politisches Verhandlungsdolmetschen. Die entscheidenden, für die Handlung und die Ziele dominanten Informationen müssen schnell und richtig umgesetzt werden. Hier kann es vorkommen, dass die Dolmetscher_innen unterbrochen werden, weshalb sich diese Situationen durch eine ganz andere Hierarchisierung von Bewertungskriterien auszeichnen (Kutz 2005:16f.).

Jede Dolmetschsituation zeichnet sich durch externe Bedingungen aus, die sowohl förderlich als auch einschränkend wirken können (Riccardi 2007:188). Diese liegen (teilweise) außerhalb des Einflussbereichs der Dolmetscher_innen. Hierzu zählen Eigenschaften der Ausgangsrede wie Geschwindigkeit, grammatische Struktur oder Inhaltsdichte. Für Dolmetscher_innen ist es von großer Bedeutung, im Vorhinein Informationen über den Kontext

und die Situation, sowie die Redner_innen und deren Absichten zu erhalten, um eine gute Dolmetschqualität zu ermöglichen. Das Profil der Rezipient_innen ist ebenfalls entscheidend (Kalina 2004:6). Auch Aspekte der Ausgangsrede spielen eine Rolle. Es macht einen Unterschied, ob eine Rede vorher schriftlich ausgearbeitet wurde und sich durch eine komplizierte Syntax auszeichnet, oder ob sie spontan vorgetragen wird (Kutz 2005:28). Garzone (2003:24) sieht Qualität als Kompromiss zwischen einem Qualitätsideal und einer realistisch realisierbaren Qualität. So ist die Qualität der Verdolmetschung immer „the result of a compromise between abstract standards and the constraints imposed by the real conditions under which it is performed“ (Garzone 2003:24f.). Zwar versuchen professionelle Dolmetscher_innen dem Ideal gerecht zu werden, müssen jedoch in einer Situation arbeiten, die mit Einschränkungen verbunden ist. Dazu zählen beispielsweise Zeitdruck und kognitive Überlastung, aber auch die Effizienz des Equipments, die Verfügbarkeit von Unterlagen oder Arbeitsbedingungen (Garzone 2003:24). Hinzu kommen auch schlechte Belüftung oder Beleuchtung der Kabinen sowie eine mangelnde Sicht (Pöhhacker 1998:329). Denn nicht immer werden die geforderten oder vertraglich vereinbarten Arbeitsbedingungen erfüllt (Pöhhacker 1994:125). Die einzelnen Faktoren der Dolmetschsituation können somit einen zusätzlichen Stressfaktor darstellen. Auf der anderen Seite dienen situative Faktoren auch als Stütze und können die Rezeption und Produktion der Dolmetscher_innen unterstützen (Pöhhacker 1998:329).

Weiters kann eine Dolmetschsituation beispielsweise als die Konstellation von zumindest drei Kommunikationspartner_innen aufgefasst werden. Dolmetscher_innen müssen die Interaktionspartner_innen antizipativ und assoziativ einschätzen (Pöhhacker 1998:328), und zwar in

Hinblick auf ihre persönlichen Merkmale, ihre person-, institutions- und situationsbezogene Rolle(n), ihr allgemeines, spezielles, individuelles und kontextuelles ‚Wissen und Können‘, ihre Befindlichkeit und ihre Wahrnehmungsbedingungen [...] sowie ihre Handlungsabsicht. (Pöhhacker 1998:328)

Dolmetscher_innen müssen das jeweilige Kommunikationsgefüge einschätzen und ihr Vorgehen anpassen, und teilweise auch zwischen dem, was allgemein gefordert ist, und jenem, was in der jeweiligen Situation machbar ist, abwägen (Mack 2002:115). Mack (2002) spricht in Zusammenhang mit Dolmetschqualität von einer Kosten-Nutzen-Relation und weist darauf hin, dass Dolmetscher_innen bereits vor dem eigentlichen Dolmetschauftrag zur Beratung herangezogen werden sollten. So könnte man die Bedürfnisse der Kund_innen und

Nutzer_innen im Vorfeld abklären und die am besten geeigneten Lösungen heranziehen. So ist neben Qualitätskontrolle am Dolmetschprodukt auch die Qualitätssicherung im Dolmetschprozess entscheidend. Die Verantwortung kann hierfür jedoch nicht allein bei den Dolmetscher_innen liegen. Für das Funktionieren müssen auch die anderen Beteiligten ihren Beitrag leisten. Dabei spielen die Bereitschaft, miteinander in Kommunikation zu treten, sich auf die anderen einzulassen und die speziellen Bedingungen einer gedolmetschten Situation miteinzubeziehen, eine zentrale Rolle (Mack 2002:117f.). Denn „[w]o diese Bereitschaft fehlt, da vermag auch der beste und bestvorbereitete Dolmetscher nur wenig auszurichten“ (Mack 2002:118). An dieser Stelle erscheint es sinnvoll, nun detaillierter auf die Bedeutung der Rezipient_innen einzugehen.

2.3.2.2. Die Rolle der Rezipient_innen

Auch wenn die Sicht der Rezipient_innen auf Qualität nicht das einzige Kriterium für die Bewertung darstellen kann, so zielt ein Text stets auf eine bestimmte Zuhörerschaft, die einbezogen werden muss (Kurz 2003:3). Kurz (2003) spricht von der „user-oriented nature“ (Kurz 2003:7) des Konferenzdolmetschens und wundert sich, dass Studien zu Erwartungen von Rezipient_innen erst so spät als Thema in der Dolmetschwissenschaft aufgegriffen wurden. Zu lange seien ihre Bedürfnisse und Erwartungen außer Acht gelassen worden. So wurde auch die erste empirische Studie zu Qualitätskriterien von Bühler (1986) mit Konferenzdolmetscher_innen anstatt mit Rezipient_innen durchgeführt (Kurz 2003:7). Bereits Seleskovitch (1986) schreibt: „[I]nterpretation should always be judged from the perspective of the listener and never as an end in itself. The chain of communication does not end in the booth!“ (Seleskovitch 1986:236) So zeichnet sich Dolmetschen stets durch einen bestimmten Bezug zu Rezipient_innen aus (Kalina 1998:272). Nach Kalina (1998) sind Rezipient_innen in einer Situation, in der authentisch gedolmetscht wird, auch überhaupt die einzigen, die basierend auf ihren Kenntnissen und Anforderungen, die Leistung überhaupt bewerten können (Kalina 1998:272).

Die Bewertungsperspektive der Zieltext-Rezipient_innen zeichnet sich jedoch durch unterschiedliche Charakteristika aus und birgt einige Variablen in sich. So interessieren sich Rezipient_innen in verschiedenem Maß für das Thema, verfügen über unterschiedliche Grade an Sachwissen und sind in Hinblick auf den Umgang mit Dolmetscher_innen in verschiedenem Ausmaß geübt. Dennoch wird an dieser Stelle, vor allem für den didaktischen Bereich, von

Prototypen gesprochen. So können Rezipient_innen kooperative Kommunikant_innen sein, die sich im Thema auskennen, an ihm interessiert sind, und Erfahrung mit dem Umgang mit Dolmetscher_innen haben. Ein zweiter Prototyp umfasst nörgelnde Personen, die nicht kooperativ sind. Weiters können Rezipient_innen charakterisiert werden, die über nicht allzu viel Sachwissen verfügen, aber am Thema interessiert sind, Erfahrungen im Umgang mit Dolmetscher_innen haben und verständnisvoll sind. Der vierte Prototyp ist das auf unangenehme Art und Weise störendes Pendant zum dritten Typen (Kutz 2005:14).

Die zahlreichen Befragungen verschiedener Rezipient_innen-Gruppen zeigten eine Reihe von unterschiedlichen Erwartungen und Prioritäten auf (Kalina 2004:5). Es hat sich auch herausgestellt, dass die Qualitätserwartungen von Rezipient_innen sowohl von der Art der Veranstaltung, also auch von individuellen Ansichten und Hintergründen abhängen (Garzone 2003:24). So können in ein- und derselben Situation verschiedene Rezipient_innen unterschiedliche Erwartungen haben (Kahane 2000:5). Während einige Personen aufgrund schlechter Erfahrungen mit unprofessionellen Dolmetscher_innen mit einer argwöhnischen Haltung in die Situation kommen, begegnen andere ihnen mit einer Art Bewunderung und haben beispielsweise herausragende Dolmetschungen im Fernsehen gesehen (Kutz 2005:28).

Die Schwierigkeit besteht aber auch darin, dass bei Untersuchungen, in denen nach der relativen Wichtigkeit von Evaluationskriterien zur Dolmetschqualität gefragt wird, keine Garantie dafür besteht, dass die Antworten tatsächlich die Kriterien widerspiegeln, die die befragten Personen bei echten Verdolmetschungen anwenden würden (Garzone 2003:25). Denn „when stating abstract views people are potentially exposed to ideological bias and distortions“ (Garzone 2003:25). So untersuchte Garzone (2003) die Übereinstimmung zwischen abstrakten, in Befragungen angegebenen Bewertungskriterien und jenen, die zur Evaluation tatsächlicher Verdolmetschungen angewendet wurden. Sie folgerte, dass wenn Qualitätsevaluation auf abstrakter Ebene besprochen wird, die Kriterien isoliert betrachtet werden. Außerdem sind diese abstrakten Vorstellungen allgemein, auf alle Situationen anwendbar formuliert und basieren zwar auf Erfahrung, aber auch auf voreingenommenen, vorgefassten Meinungen. Bei einer tatsächlichen Evaluierung überlappen sich die Kriterien und beeinflussen einander (Garzone 2003:28).

Rezipient_innen spielen an dieser Stelle jedoch eine zentrale Rolle, denn „[i]t is true that the main criterion for the evaluation of an interpreting performance is its impact on the listener“ (Riccardi 2002:116). Eines der Probleme der Qualitätsforschung ist jedoch, dass Rezipient_innen oftmals keine Möglichkeit haben, den ZT mit dem AT zu vergleichen. Sie müssen ihr Urteil somit auf indirekte Indikatoren stützen, das Gehörte mit ihrem Wissen

abgleichen und in Hinblick auf Plausibilität und Kohärenz betrachten (Garzone 2003:28). Seitens der Rezipient_innen können also sehr wohl unsinnige oder widersprüchliche Äußerungen, unvollendete Sätze, sowie Schwierigkeiten mit der Fachterminologie und bei der Wiedergabe erkannt werden (Riccardi 2002:116). Auch aus einer soziologischen und kognitionspsychologischen Perspektive nehmen Rezipient_innen in einer Interaktion die wichtigste Rolle ein: Sie weisen dem Gesagten den Sinn zu (Behr 2013a:34). „Da Verstehen, Sinn und Bedeutungszuweisung [...] letztlich allein beim original- wie fremdsprachlichen Zuhörer liegen, sollte das Augenmerk stärker auf ihn gelenkt werden.“ (Behr 2013a:35) Die Verarbeitung des Textes seitens der Rezipient_innen hängt jedoch von zahlreichen Variablen ab, die nicht alle im Einflussbereich der Dolmetscher_innen liegen (Kalina 2004:5).

Behr (2013a) schenkt in ihrer Arbeit (siehe hierzu Kapitel 4.3.2.) den Rezipient_innen besondere Aufmerksamkeit und rückt das Verstehen als zentralen Bestandteil einer Kommunikations- bzw. Dolmetschsituation in den Vordergrund. Dabei spielt es auch eine Rolle, „inwieweit – aus sozialpsychologischer Sicht – ein Rezipient verstehen kann und will“ (Behr 2013a:19). Hierbei sind sowohl Erwartungshaltungen, Einstellungen, Werte, Normen, Motivation und schließlich die Befindlichkeit von Bedeutung. Sie rückt dabei insbesondere die Stimmung (siehe hierzu Kapitel 4.1.1.) ins Zentrum (Behr 2013a). Behr (2013b:88-97) bezieht in Hinblick auf die Evaluation im Dolmetschen Erkenntnisse aus den Kommunikationswissenschaften und der Systemtheorie nach Luhmann (1991) ein. Sie spricht von einem „Gefüge der Evaluation“ (Behr 2013b:88), das aus den drei Komponenten Setting, Kriterienkonsens und Disposition besteht (Behr 2013b:96). Um der Befindlichkeit bzw. Stimmung der Zuhörer_innen auf den Grund zu gehen, muss ein Bogen vom Bereich der Kommunikationswissenschaft hin zur Soziologie, oder konkreter in die sozialpsychologische Stimmungsforschung, gespannt werden. Denn bei der Erforschung der Wirkungsäquivalenz bei den Zuhörer_innen muss der Einfluss der Stimmung auf die Rezeption einer Person betrachtet werden (Behr 2013a:19). Die vorliegende Arbeit folgt diesem Ansatz und stellt die Rezipient_innen und deren Stimmung bei der Evaluation in den Vordergrund. Kapitel 4 ist daher den Affekten gewidmet. Da die vorliegende Arbeit den Stimmungseinfluss auf Bewertungen durch Studienkolleg_innen untersucht, soll als nächstes auf die Evaluation im didaktischen Setting eingegangen werden.

2.3.3. Qualitätsbewertung im didaktischen Setting

Obwohl die Praxis des Dolmetschens bis in die Antike zurückreicht, wurde ihr bis zum 20. Jahrhundert kaum besondere Aufmerksamkeit zuteil. Mit der Professionalisierung zu Beginn des Jahrhunderts wurden in Europa jedoch nach und nach Ausbildungsmöglichkeiten für Dolmetscher_innen auf Universitätsniveau eingerichtet (Pöchhacker 2004:27f.). Die Didaktik des Dolmetschens umfasst sowohl das Simultan- als auch Konsekutivdolmetschen, sowie die Stegreifübersetzung. Hinzu kommen der Bereich Berufsethik und das Erlernen der Methodik zur Einarbeitung in Thematik und Terminologie eines Fachgebiets (Déjean Le Féal 1998:363ff.). Evaluation spielt im didaktischen Bereich eine zentrale Rolle (Sawyer 2004:5), denn „[h]igh quality education is based upon sound assessment“ (Sawyer 2004:5). Außerdem stellt die Thematik der Qualitätsbewertung im didaktischen Bereich eine wichtige Schlüsselfrage dar (Kutz 2005:14), denn sie „offenbart [...] in der Lehre, ob die Ausbildung an einer Einrichtung überhaupt systematisch, nach einem kohärenten und effektiven pädagogischen Ansatz erfolgt“ (Kutz 2005:14). Auch wenn bei Bewertungen von Studierenden nicht direkt auf Qualität Bezug genommen wird, so wird implizit nach ihr gesucht (Moser-Mercer 1996:52).

Wie in Kapitel 2.2.2. erwähnt, kann der Bereich der Dolmetschdidaktik als soziales System aufgefasst werden, das seine Auffassung von Qualität konstruiert. Bedenkt man den relativen Charakter von Qualität, so sind auch im Hochschulbereich unterschiedliche Interessent_innen auszumachen, die jeweils ein unterschiedliches Interesse und eine spezifische Sichtweise auf Qualität haben. Dazu gehören Student_innen und Lehrende, aber auch Arbeitgeber_innen und Administrator_innen oder Geldgeber_innen und die Regierung (Harvey & Green 2000:17). Kutz (2005) nennt zusätzlich zu den Bewertungsperspektiven der Konferenzveranstalter_innen und Rezipient_innen des Zieltextes drei weitere Perspektiven, die im didaktischen Bereich bedeutend sind. Hierzu zählen zunächst die Ausbilder_innen und ihre didaktisch geprägte Sicht. Ziel dieser Personen ist eine optimale und professionelle Kompetenz der Studierenden. Daher kann diese Perspektive als problem- und prozessbewusster charakterisiert werden. Die zweite Perspektive ist die der Studierenden selbst. Sie nehmen die Position von Zuschauer_innen ein und vereinen die Sichtweise von Lerner_innen sowie Beobachter_innen einer Dolmetschhandlung. Als letztes ist an dieser Stelle auch die Perspektive der Dolmetschwissenschaftler_innen zu nennen. Je nach Interesse können für diese Personen unterschiedliche Aspekte von Bedeutung sein (Kutz 2005:15). Auch Kalina (1998) spricht von verschiedenen Bewertungsperspektiven im didaktischen Bereich und nennt

Rezipient_innen-Orientierung, Lernzielorientierung, Prozessorientierung und Lernfortschrittorientierung als Beispiele (Kalina 1998:272). So zeichnet sich die Perspektive auf Qualität dadurch aus, dass diese anhand des Erreichens von Teilzielen bewertet werden kann. Lehrende sind in den Entwicklungsprozess zu der später geforderten Qualität eingebunden und können die unterschiedlichen Punkte des Lernprozesses berücksichtigen (Gile 2003:110). Die Bewertung in der Didaktik „differs from assessment in the field due to its guiding function” (Gile 2001:379). Schließlich kann eine Bewertung in diesem Setting auch einfach dazu dienen, im Rahmen einer Prüfung eine Note zu vergeben (Bartłomiejczyk 2007:248).

In Bezug auf Dolmetschen als Produkt oder als Prozess hält Sawyer (2004) fest: „Generally, though, quality refers to the features of a performance as product, that is, whether an interpretation is accurate, complete, stylistically appropriate, etc.” (Sawyer 2004:93f.) Für den didaktischen Bereich heißt dies, „it must be determined whether a student has acquired the ability to *produce* an acceptable product, or to *perform* acceptably” (Sawyer 2004:94). Dies sei mit *performance assessment* möglich, sprich durch den Fokus auf den Prozess. Hierbei müssen Studierende also ihre erreichten Fertigkeiten und ihr Verständnis unter Beweis stellen, indem sie eine oder mehrere Aufgaben durchführen, die in der realen Welt existieren. Dies kann beispielsweise das Dolmetschen einer Konferenzrede sein (Sawyer 2004:93f.). Denn

[b]ei der Evaluierung ist entscheidend, daß [sic] nicht nur das Produkt, also die Verdolmetschung selbst, in die Betrachtung einbezogen wird, wie dies aus der Sicht des Rezipienten in der Regel geschieht, sondern daß [sic] prozeßorientierte [sic] Aspekte berücksichtigt werden. (Kalina 1998:273)

Während der Fokus auf den Prozess vor allem zu Beginn der Ausbildung zentral ist, herrscht ein Konsens darüber, dass der Fokus im Verlauf bzw. gegen Ende der Ausbildung auf das Produkt verlagert werden sollte (Dastyar 2019:21). Dennoch gilt, „we must remember that errors in the product are often indicative of errors in the process and it is often not easy to separate the two” (Martí & Martin 2003:204).

2.3.3.1. Bewertungstypologien in der Dolmetschdidaktik

Im didaktischen Setting des Dolmetschens existieren unterschiedliche Arten bzw. Typologien von Bewertungen. Diese können anhand von drei Merkmalen festgemacht werden. Zum einen kann danach unterschieden werden, zu welchem Zweck eine Bewertung vorgenommen wird. Hierbei wird von ipsativer, formativer oder summativer Bewertung gesprochen. Zweitens kann

der Bewertungstyp anhand des Zeitpunktes in der Ausbildung festgemacht werden. Schließlich kann eine Unterscheidung auch danach getroffen werden, wer eine Bewertung vornimmt (Dastyar 2019:24ff.).

Zu Beginn einer Ausbildung werden Bewertungen beispielsweise im Rahmen eines Auswahlverfahrens durchgeführt. Sie können aber auch als eine Art Bestandsaufnahme gelten, die für den weiteren Ausbildungsweg von Nutzen sein kann. Diese Bewertungen beziehen sich jedoch eher auf Aspekte wie Allgemeinbildung oder Sprachenkenntnisse und nicht auf das Dolmetschen (Gile 2001:381). Solche Eignungstests sollen vor Beginn der Ausbildung Aufschluss darüber geben, ob Kandidat_innen die für notwendig erachteten Grundvoraussetzungen besitzen (Martí & Martin 2003:197).

Die Evaluation im Verlauf der Ausbildung bietet den angehenden Dolmetscher_innen eine Orientierung und zeigt ihnen, wo sie sich befinden (Gile 2001:381). Bei Bewertungen in den Lehrveranstaltungen, die ein Feedback zum Lernprozess darstellen, spricht man von formativer Bewertung. Hier fallen auch Beurteilungen von Aufgaben und Feedback zu Kursarbeiten oder Selbsteinschätzungen hinein (Sawyer 2004:107). Dabei müssen nicht zwingend Noten vergeben werden (Dastyar 2019:25). Im Ausbildungsverlauf kann es jedoch auch zu abschließenden Prüfungen kommen, die das Erreichen der allgemeinen Ziele der Lehrveranstaltungen erfassen sollen (Martí & Martin 2003:201).

Schließlich kommt es auch am Ende der Ausbildung zu einer Reihe von Prüfungen. Diese geben Aufschluss darüber, ob die Studierenden prinzipiell für den Arbeitsmarkt geeignet sind. Dabei werden sie allgemein nach demselben Maßstab beurteilt, der auch für professionelle Dolmetscher_innen gilt (Gile 2001:381). Diese Bewertungen fallen in die Kategorie der summativen Bewertung und werden auch als *terminal assessment* bezeichnet. Sie finden zu bestimmten Zeitpunkten statt und können beispielsweise mit der Vergabe einer Note zusammenhängen (Dastyar 2019:26). Hierbei soll sowohl überprüft werden, wie effektiv der Lehrprozess war, als auch wie gut die Studierenden gelernt haben (Sawyer 2004:107).

Bei der ipsativen Evaluation handelt es sich schließlich um eine Selbstevaluation, die durch eine reflektierende Person durchgeführt wird. Dabei wird die aktuelle Leistung entweder mit den eigenen, vorangegangenen Leistungen, oder mit jenen anderer verglichen. Diese Form stellt eine fortlaufende Reflexion dar, die idealerweise kontinuierlich während der professionellen Laufbahn durchgeführt wird. Die ipsative Bewertung integriert auch Rückmeldungen von anderen, sprich den Lehrenden und den Kolleg_innen (Sawyer 2004:107).

Die Bewertung im didaktischen Bereich kann daher weiters danach unterteilt werden, ob es sich um eine Selbstevaluation oder eine externe Bewertung handelt (Bartłomiejczyk

2007:251). Die Beurteilung kann von Seiten der Lehrenden, der Studienkolleg_innen oder durch die Person, dessen Dolmetschung zu evaluieren ist, erfolgen (Kalina 1998:272f.). Dabei wird das Thema der Selbstevaluation bzw. die Frage, wie Dolmetscher_innen selbst ihren Output beurteilen, in der Literatur kaum behandelt (Bartłomiejczyk 2007:247). Bartłomiejczyk (2007) setzt an dieser Stelle an und beschäftigte sich mit Reaktionen von fortgeschrittenen Studierenden auf ihre Leistungen. Dabei konnte eine negative Tendenz in Hinblick auf die Bewertung beobachtet werden. Zudem wurde Aspekten wie der Treue zum Original und der Vollständigkeit die meiste Aufmerksamkeit geschenkt, während Aspekte wie Intonation, Stimme und Pausen kaum angesprochen wurden (Bartłomiejczyk 2007:247). Ihre empirische Untersuchung basiert auf Erkenntnissen einer ihrer vorangegangenen Arbeiten, in der eine negative Tendenz bei den Kommentaren der Studierenden und eine kritische Haltung beobachtet werden konnten. In einem neuen Versuch, in dem die Teilnehmer_innen durch Retrospektive und anhand eines eigens angefertigten Transkripts ihre Leistung kommentieren sollten, konnte beobachtet werden, dass den Studierenden die Beurteilung des Dolmetschproduktes leichter fiel als die Retrospektion zu strategischen Prozessen, die eher unbewusst zu sein scheinen. Zwar konnten weniger negative Kommentare beobachtet werden, dennoch waren mehr als die Hälfte der analysierten Segmente negativer Natur. Die Ergebnisse ihrer Untersuchungen zeigten, dass Selbstevaluation mit Fokus auf negative Aspekte seitens der Studierenden als natürliche Tätigkeit angesehen werden kann. Dennoch bleibt die Frage offen, welche Aspekte der Dolmetschqualität durch Selbstevaluation verbessert werden können (Bartłomiejczyk 2007:254-264).

Auch für *peer evaluation*, auch *peer assessment* oder *peer feedback* genannt, fehlt es an ausreichend empirischen Studien (Dastyar 2019:29f.). Dennoch weist Kalina (1998) darauf hin, dass in der Literatur allgemein die *peer evaluation* empfohlen wird, bevor die Lehrenden genauer erläutern und ergänzen. Hierbei beurteilen zunächst Studienkolleg_innen eine Dolmetschung. Aber auch in Übungsgruppen ohne Lehrende spielt diese Art der Bewertung eine große Rolle. Zu den Vorteilen dieser Methode zählt, dass Kritik weniger persönlich genommen wird, wenn sie von gleichaltrigen Personen geäußert wird, was wiederum zu einer höheren Akzeptanz führt. Zudem lernen Studierende hierbei, die Dolmetschleistungen anderer auf angemessene Art und Weise und mit den geeigneten Worten zu evaluieren. Schließlich wird durch die *peer evaluation* auch geübt, im Team zu arbeiten, sich mit den Aufgaben abzuwechseln, sich gegenseitig zu unterstützen und Kritik zu üben (Kalina 1998:272). Ein *peer feedback* kann beispielsweise anhand von Kategorien oder Checklisten erfolgen und sowohl im Rahmen einer formativen als auch summativen Bewertung vorgenommen werden. Dabei ist es

wichtig, dass die Studierenden entsprechend darauf vorbereitet werden, anderen Feedback zu geben (Dastyar 2019:29). Je fortgeschrittener die Student_innen, desto besser scheint die *peer evaluation* geeignet zu sein. Wird sie zu früh eingesetzt, kann sie den beurteilten Personen womöglich ein falsches Bild ihrer Leistung vermitteln, da Studierende dazu neigen, die Dolmetschungen ihrer Kolleg_innen sehr mild zu bewerten, wenn die Lehrenden anwesend sind. Auch die Bewertung der eigenen Leistung wird mit dem Studienfortschritt ausgewogener, analytischer und sicherer. Dies liegt daran, dass den Studierenden im Laufe der Zeit die verschiedenen Dimensionen des Dolmetschens bewusster werden und sie auch in der Lage sind, ihre Kapazitäten besser aufzuteilen (Kalina 1998:272f.).

So scheinen je nach Fortschritt der Studierenden unterschiedliche Vorgehensweisen bei der Bewertung geeignet zu sein. Anfangs müssen die Lernziele klar und deutlich erläutert werden. Die Lehrenden übernehmen die Lernkontrolle (Kalina 1998:273). Für diese anfängliche Phase wird eine Didaktik empfohlen, die als positiv stimulierend bezeichnet werden kann. Hierbei sollen mangelhafte Aspekte in der Leistung entdramatisiert werden und den Studierenden hilfreiche Hinweise geboten werden, während positive Leistungen hervorgehoben werden (Kutz 2005:16). Nach einem gewissen Fortschritt erfolgt die Bewertung aus drei Perspektiven: Selbstevaluation, Beurteilung durch Lehrende und die Bewertung durch Kolleg_innen (Kalina 1998:273) und „[s]o verschiebt sich die Evaluierungsperspektive langsam hin zu einer immer stärker rezipientengeprägten Beurteilung“ (Kalina 1998:273).

Das Feedback der Lehrenden ist auch deshalb von großer Bedeutung, da es auch einen Einfluss auf die Selbsteinschätzung der Studierenden zu haben scheint (Bartłomiejczyk 2007:263), denn „[the] students‘ spontaneous comments on their own output reflect to a great extent the feedback they usually receive in class from their trainer“ (Bartłomiejczyk 2007:263). Es gilt als allgemein anerkannt, dass für die Bewertung von Studierenden eine Auflistung von Fehlern oder unglücklich gewählten Formulierungen beim Dolmetschen nicht ausreichend ist. Stattdessen sollte thematisiert werden, was hinter den Mängeln liegt, sprich wo und warum die Leistung der Studierenden einbricht und wo angesetzt werden kann, um dies in Zukunft zu vermeiden (Sunnari 2003:235). Bei der Bewertung wird der Fokus jedoch aufgrund der beschränkten Zeit in den Lehrveranstaltungen, vor allem der beschränkten Zeit in der Kabine, oft auf negative Aspekte der Leistung und auf die größten Fehler gelegt, während Aspekte hinsichtlich der Präsentation vernachlässigt werden (Bartłomiejczyk 2007:263). Bartłomiejczyk (2007) hält hierzu aus der Sicht der Lehrenden fest: „Therefore, as responsible trainers we need to diversify our feedback to include all aspects of interpreter performance and to comment more often on the successful solutions we observe in our students‘ outputs.“

(Bartłomiejczyk 2007:263) Doch auch für Lehrende ist es nicht immer einfach, die Leistungen ihrer Student_innen zu erfassen und zu erklären. Zum Teil scheint dies am Fehlen einer geeigneten Metasprache für die Bewertung von Dolmetschleistungen in diesem Bereich zu liegen (Moser-Mercer 2000:339). Daher muss auf einer allgemeineren Ebene angesetzt werden, um dort die Schwierigkeiten des Konferenzdolmetschens herauszuarbeiten. So kann den Wurzeln der Probleme auf den Grund gegangen werden, um anschließend passendes Feedback geben zu können. Einen möglichen Ansatz stellt an dieser Stelle der Vergleich von Expert_innen und Noviz_innen dar. Diese Perspektive führt in den Bereich der Expertiseforschung (Sunnari 2003:235f.), die in Kapitel 3 behandelt wird. Doch zuvor soll ein Blick in die Evaluationskriterien in der Ausbildung geworfen werden.

2.3.3.2. Evaluationskriterien

In Hinblick auf Bewertungen in Form von Eignungstests sind in der Dolmetschliteratur zahlreiche Arbeiten vorhanden. Dies ist jedoch nicht der Fall für die anderen Formen der Bewertung im didaktischen Bereich (Martí & Martin 2003:197). Somit besteht „a vacuum as far as criteria for continuous assessment and final exams are concerned“ (Martí & Martin 2003:197). Für die Studierenden ist es jedoch von großer Bedeutung, über die beispielsweise bei Prüfungen verwendeten Kriterien im Vorhinein aufgeklärt zu werden, da dies ihre Eigenständigkeit im Lernprozess verbessert und auch einen bedeutenden Einfluss auf die Qualität ihrer Leistungen haben kann. So entwickelten einige erfahrene Lehrende für den didaktischen Bereich Bewertungsbögen inklusive Bewertungskriterien. Diese sollen dazu dienen, die Leistungen der Studierenden zu analysieren und zu beurteilen (Bartłomiejczyk 2007:251).

Wie bereits in Kapitel 2.3.1. beschrieben, entwickelte beispielsweise Riccardi (2005) Makrokriterien zur Bewertung von Simultandolmetschleistungen. Solche Makrokriterien eignen sich für die Evaluation von professionellen Verdolmetschungen (Riccardi 2002:118), während für Bewertungen von Student_innen die Mikrokriterien herangezogen werden können (Riccardi 2002:118; Riccardi 2005:198). In einem anderen Beitrag beschreibt Riccardi (2007) die Mikrokriterien als eine Liste von 14 Kriterien, die sowohl für Bewertungen im Simultan- als auch Konsekutivdolmetschen eingesetzt werden können. In dieser Auflistung ist beispielsweise das Décalage nicht mehr als eigener Punkt aufgelistet, und das Mikrokriterium zur Bewertung der Gesamtleistung wird als „Funktionalität“ bezeichnet. Diese bezieht sich auf eine ganzheitliche Evaluation bzw. eine globale Bewertung und soll anzeigen, ob die Funktion

der Verdolmetschung für die Zuhörer_innen erfüllt wurde. Für die beiden Modi kommen jeweils spezifische Kriterien dazu. Die Bewertung von Konsekutivleistungen wird durch das Kriterium „Augenkontakt“ sowie „Kontrolle der Hände“ erweitert, das Simultandolmetschen um das Mikrokriterium „nicht beendete Sätze“ (Riccardi 2007:199f.). Riccardi (2007) weist darauf hin, dass Unterschiede in der relativen Gewichtung der einzelnen Bewertungskriterien zu den unterschiedlichen Zeitpunkten im Ausbildungsverlauf festgestellt werden konnten (Riccardi 2007:198).

Auch Kutz (2005) entwickelte ein Beurteilungsschema für den didaktischen Bereich, wobei er die zwei Modi des Konsekutiv- und Simultandolmetschens getrennt behandelt. Er spricht von einer Bewertungshandlung, die sich aus zwei Teilen zusammensetzt. Das Ziel der aufgestellten Bewertungskriterien ist es, alle Aspekte, die potenziell relevant sein könnten, zu erfassen und somit einen handlichen Begriffsapparat zur Verfügung zu stellen, der für eine Bandbreite an möglichen Fällen eingesetzt werden kann. Dabei gilt, dass in einem konkreten Bewertungsfall bzw. in einer konkreten Bewertungshandlung nicht alle Einzelaspekte angesprochen bzw. analysiert werden müssen. Im ersten Teil (Teil A) erfolgt eine Bewertung der Verdolmetschung nach den drei Grundaspekten „Eindruck“, „sprachliche Form“ und „Inhalt“, wobei ihre Rangfolge nicht objektiv fest steht, sondern davon abhängt, welcher Wert den Aspekten im Rahmen eines Dolmetschauftrags beigemessen wird. Dieser erste Teil stellt also eine Art Diagnose dar und endet mit einer Zusammenfassung. Im zweiten Teil (Teil B) werden basierend auf den festgestellten Mängeln oder erkannten Reserven didaktische Schlüsse gezogen. Dieser Teil kann also als eine Art Therapie bezeichnet werden. Prinzipiell kann eine Bewertung nach diesem Prinzip verschieden gestaltet sein und auch durchaus knapp ausfallen (Kutz 2005:20-31).

Für die Bewertung von Simultandolmetschungen stellt Kutz (2005) zwei Tabellen vor, wobei eine für die Bewertung im didaktischen Setting angewendet werden kann und die Perspektive der Ausbilder_innen einschließt, während die zweite die Perspektive der zielsprachlichen Zuhörer_innen bzw. der Kund_innen aus der Praxis umfasst. Beide Bewertungsmodi setzen sich aus zunächst einem ersten Teil, der Leistungsanalyse, zusammen, die die drei Aspekte „Höreindruck“, „Inhaltswiedergabe“ und „sprachliche Realisierung“ umfasst. Dem folgen, wie bereits angemerkt, eine Zusammenfassung und Schlussfolgerungen. Für die Kund_innen stellt eine Zusammenfassung ein Gesamturteil dar, das ihnen ein Gefühl dafür gibt, was bei einer Dolmetschleistung möglich ist und wo die Grenzen liegen. Schlussfolgerungen beziehen sich in diesem Fall darauf, welche Ansprüche an Dolmetscher_innen gestellt werden können oder welchen Dolmetscher_innen man in Zukunft

vertrauen kann. In der Dolmetschdidaktik handelt es sich beim zweiten Teil um Schlussfolgerungen didaktischer Art, wobei basierend auf der Zusammenfassung der Analyse aus dem ersten Teil konkrete Übungen vorgeschlagen werden (Kutz 2005:30f.).

Unter „Höreindruck“ fällt nach Kutz (2005) zunächst eine funktionale Analyse der Stimme bzw. Tonlage hinein. Hier sollen zum Beispiel die Fragen beantwortet werden, ob diese für die Situation und den Auftrag der Dolmetschung angemessen waren oder ob sie dem Original ähnlich waren. Dem folgt der Punkt „Psychologie der Stimme“, in dem beispielsweise die Stimme in Hinblick auf Müdigkeit oder die Stimmhöhe untersucht wird. Der dritte Punkt im Höreindruck bezieht sich auf die Prosodie und umfasst u.a. das deutliche Artikulieren, Betonungen, den Intonationsverlauf oder die Lautstärke. Der nächste Unterpunkt umfasst, was Kutz (2005) als „stimmliche Anmutsähnlichkeit“ mit den Redner_innen nennt. Diese bezieht sich beispielsweise darauf, ob die Person der Sprecher_innen durchscheint oder ob deren narrative Rolle mitgespielt wird. Die letzten Punkte des „Höreindrucks“ in diesem Bewertungsschema sind die Selbstbeherrschung der Dolmetscher_innen bzw. ihre Emotionen, die Atemtechnik und alle möglichen Geräusche, die stören (Kutz 2005:30).

Die weiteren Blöcke umfassen die Aspekte „Inhaltliche Konstanz“ und „Sprachliche Realisierung“. Ersterer bezieht sich auf die Wiedergabe des Inhalts und der unterschiedlichen Informationen. Hier soll beispielsweise das professionelle Wiedergeben von grundlegenden Sachverhalten oder die Schilderung in ihrer Kohärenz beurteilt werden. Auch der letzte Block („Sprachliche Realisierung“) beinhaltet drei Unterpunkte, die wiederum einige Aspekte umfassen. Im ersten Unterpunkt wird die Umsetzungsweise von sowohl strategischen Aspekten als auch Dolmetschtechniken analysiert. Die dolmetschstrategischen Aspekte beziehen sich beispielsweise darauf, wie mit sprachkulturellen Unterschieden oder Mängeln in der Originalrede umgegangen wird, oder wie Unterschiede im Redestil, die kulturgeprägt sind, bewältigt wurden. Die Techniken werden in syntaxbezogene und lexikalisch bezogenen Dolmetschtechniken unterteilt. Der nächste Unterpunkt bezieht sich auf den Redefluss und analysiert zum Beispiel dessen Natürlichkeit. Der letzte Unterpunkt wird als „Systemsprachliches“ bezeichnet und umfasst die Grammatik und den Stil, sowie die Betonung und den phonomatischen Aspekt der Aussprache (Kutz 2005:30f.).

Für Prüfungen oder für Bewertungen im Rahmen von Übungen werden Bögen eingesetzt, die Kategorien zur Präsentation, zur Sprache, zum Inhalt und zu den eingesetzten Strategien umfassen (Riccardi 2007:201). Am Zentrum für Translationswissenschaft in Wien wird für die abschließenden Modulprüfungen des Masterstudiums mit dem Schwerpunkt Konferenzdolmetschen für die Simultandolmetschungen ein Bewertungsbogen mit drei

Kriterienblöcken verwendet. Diese sind „Inhalt“, „Zielsprache“ und „Dolmetsch-Fertigkeiten & Präsentation“, die jeweils 5 bis 7 Unterpunkte enthalten (ZTW o.J.).

Nachdem nun die wichtigsten Aspekte des Konzeptes Qualität für diese Arbeit erläutert wurden, soll im nächsten Kapitel auf die Expertise eingegangen werden.

3. Expertise

Bis ins 20. Jahrhundert wurde dem Dolmetschen weder ein bestimmter professioneller Status noch eine eindeutige Berufsdefinition zugeschrieben (Pöchhacker 2004:27f.). Doch spätestens nach dem Zweiten Weltkrieg galt das Konferenzdolmetschen als etablierter Beruf (Pöchhacker 2013:37). Diese Anerkennung „of interpreting as a profession implies that there is a body of specialized knowledge and skills“ (Pöchhacker 2004:32). Diese professionelle Expertise wird aufgrund von Erfahrungen und Reflexion entwickelt (Pöchhacker 2004:32).

In ihrer Arbeit weist Behr (2013a) darauf hin, dass die Entwicklung von Dolmetschexpertise im Rahmen der Ausbildung womöglich eine Rolle dabei spielen könnte, ob sich Dolmetscher_innen beim Bewerten von ihren Stimmungen leiten lassen (Behr 2013a:301f.). Da hierauf aufbauend die Expertise auch für das Forschungsinteresse der vorliegenden Arbeit zentral ist, erscheint es sinnvoll, sich auch mit diesem Konzept näher auseinanderzusetzen. Im Folgenden soll zunächst auf die Frage eingegangen werden, was genau unter Expertise zu verstehen ist und welche Faktoren eine Rolle für ihre Entwicklung spielen. Dem folgt eine Auseinandersetzung mit Expertise im Bereich Dolmetschen.

3.1. Zur Definition von Expertise

Unabhängig vom Fachgebiet gilt Expertise als hart erarbeitete, wertvolle Ressource (Hoffmann 1997:191). Die Expertiseforschung als Teil der kognitiven Psychologie (Pöchhacker 2004:167) untersucht Höchstleistungen, die Personen auf ihrem Spezialgebiet erbringen (Winter 2011:210). Im Fokus stehen dabei die Erklärung solcher Leistungen, das Voraussagen des Entwicklungspotenzials einer Person, sowie die Schaffung von bestmöglichen Voraussetzungen für den Erwerb von Expertise (Winter 2011:210). Dabei untersucht die Expertiseforschung stets ein „Produkt“, sprich eine erbrachte Leistung. Als Ausgangspunkt gilt bereits ein Zielzustand, nämlich die Meisterung des zu untersuchenden Wissensgebiets (Chabasse 2009:46f.).

Die moderne Expertiseforschung reicht bis in die 1960er Jahre zurück (Sunnari 2003:236). Zu den klassischen untersuchten Bereichen zählen Schachspielen, Musik, Medizin und Sport (Ericsson 2000:189). Zunächst wurde Expertise anhand von klar definierbaren Aufgaben, die typisch für einen Bereich waren, untersucht und als individuelle Fertigkeit angesehen. In einem anderen Ansatz wird Expertise als soziale Handlung angesehen. Ihre Entwicklung erfolgt in der Interaktion mit Expert_innen. Ein dritter, deskriptiver Ansatz stellt

den Vergleich von Expert_innen und Noviz_innen in den Fokus und thematisiert die Entwicklung zum Expert_innen-Status. Nach diesem Ansatz wird Expertise zumeist als außergewöhnlich gute Problemlösungskompetenz beschrieben. Unter diesen Ansatz fällt ein Großteil der Forschung zur Expertise (Sunnari 2003:236). Dabei wird davon ausgegangen, dass besonders gute Leistungen bei Expert_innen durch das bereichsspezifische Wissen erklärt werden können. Auch in der Dolmetschforschung zur Expertise (siehe Kapitel 3.2.) wurde der Expert_innen-Noviz_innen-Vergleich aufgegriffen (Chabasse 2009:53ff.).

An dieser Stelle stellt sich die Frage, was Expert_innen konkret ausmacht. Basierend auf Studien in unterschiedlichen Bereichen konnten einige allgemeine, auf verschiedene Gebiete anwendbare Eigenschaften gefunden werden. So zeichnen sich Expert_innen dadurch aus, dass sie unterschiedliche Tätigkeiten parallel bewältigen können. Außerdem verfügen sie über ein breites Wissen über ihr Fachgebiet, können bessere Vorhersagen treffen als Noviz_innen und besser mit neuen Situationen umgehen. Expert_innen verfügen über eine geschulte Wahrnehmung, wodurch sie komplexe Informationen schnell verarbeiten, Ursachen und Ziele schneller erkennen und darauf basierend schnellere Entscheidungen treffen können. Neue Informationen werden auch leichter erfasst und gespeichert als bei Anfänger_innen. Gleichzeitig wird gespeichertes Wissen effektiver abgerufen. Weiters führen Expert_innen ihre Tätigkeit schneller aus und lösen Probleme mit wenigen Fehlern. Sie überwachen sich selbst und können auch fehlende Informationen bzw. Elemente erschließen oder aus gespeicherten Informationen konstruieren. All diese Eigenschaften sind auch für die Expertise im Dolmetschen (siehe hierzu Kapitel 3.2.) von Vorteil, und Personen, die über diese Fähigkeiten verfügen, werden höchstwahrscheinlich bessere Leistungen im Dolmetschen erbringen (Sunnari 2003:236f.).

Um zu Expertise zu gelangen, spielt Erfahrung eine zentrale Rolle (Casado et al. 2000:108; Ericsson 2000:189; Hoffmann 1997:216; Moser-Mercer 2000:339; Pöchlacker 2004:32). So definiert Moser-Mercer (2000) Expert_innen generell als „someone who has attained a high level of performance in a given domain as a result of years of experience“ (Moser-Mercer 2000:339). Doch Erfahrung allein genügt nicht, um wahre Expertise auf einem Gebiet zu erlangen (Chabasse 2009:47; Ericsson 2000:197f.). Denn sobald ein gewisses Niveau erreicht ist und eine Automatisierung eintritt, reicht weitere Erfahrung allein nicht mehr aus, um die Leistung zu steigern und zu verbessern (Ericsson 2000:197f.). Hinzu kommt die Notwendigkeit des Übens (Chabasse 2009:48f.; Hoffmann 1997:216). Doch auch die Begabung bzw. die Kompetenz einer Person spielt für den Expertiseerwerb eine Rolle. So muss die

Beschäftigung mit einem Gebiet nicht zwingend dazu führen, dass eine Person zum Experten bzw. zur Expertin wird (Chabasse 2009:47).

So soll im Folgenden auf Erkenntnisse der Kognitionswissenschaft und Begabungsforschung eingegangen werden, die für die Entwicklung von Expertise eine Rolle spielen (Winter 2011:203).

3.1.1. Erkenntnisse aus der Kognitionswissenschaft und der Begabungsforschung

Wie bereits angedeutet, ist Problemlösen ein entscheidender Aspekt für Expertise. Unter diesen Punkt fallen ebenfalls der Wissenstransfer und das prozedurale Wissen. Zunächst kann Problemlösen als Denken und Handeln definiert werden, das zielorientiert ist und durch Anwendung von Lösungsschritten zur Erreichung von Teilzielen führt. Dabei hängt es von der Erfahrung der Person ab, ob es sich prinzipiell um ein Problem handelt. Man unterscheidet zwischen *well-defined* und *ill-defined problems*. Erstere stellen klar definierte Probleme dar, bei denen ein Endzustand klar formuliert und auch überprüfbar ist. *Ill-defined problems* wiederum haben meist keinen eindeutig definierten Zielzustand und sind nicht zweifelsfrei überprüfbar. Des Weiteren wird zwischen *knowledge-lean* und *knowledge-rich problems* unterschieden. Im Gegensatz zur ersten Gruppe verlangen *knowledge-rich problems* für die Lösung ein besonderes Wissen. Tätigkeiten, die zu dieser Gruppe zählen, bedürfen meistens einer speziellen Ausbildung (Winter 2011:203ff.).

Für das Lösen von Problemen ist die Anwendung bereits erprobter Lösungsstrategien auf andere Problemstellungen, sprich der Wissenstransfer, von Bedeutung. Dabei wird zwischen nahem Wissenstransfer für Probleme in ähnlichen Kontexten und dem weiten Wissenstransfer unterschieden. Letzterer setzt voraus, dass das Problem aus einer neuen Perspektive betrachtet und umgedeutet wird, um Analogien zu bereits zuvor gelösten Problemen zu erkennen. Diese Fähigkeit scheint bei vielen Personen jedoch schwächer vorhanden zu sein als angenommen. Auch bei Menschen mit langer Erfahrung auf einem Gebiet kommt es vor, dass sie ihre Fähigkeiten nicht auf benachbarte Disziplinen anwenden können. So scheint auch die Intelligenz eine Rolle für den Wissenstransfer zu spielen (Winter 2011:205ff.). Inwiefern der Wissenstransfer eine Rolle für Expertise spielt, scheint in der Forschungsgemeinschaft jedoch umstritten zu sein (Winter 2011:210). Winter (2011:210) weist hier auf zwei entgegengesetzte Standpunkte hin: Während Setton (2007) die Fähigkeit, sich einer Aufgabe anzupassen und unterschiedliche Fähigkeiten entsprechend zu kombinieren, als

Teil der Expertise nennt (Setton 2007:65), betont Ericsson (2000), dass einige Expert_innen nicht dazu in der Lage sind, modifizierte Aufgaben bzw. Aufgaben außerhalb ihres Fachgebiets gut zu bewältigen (Ericsson 2000:208).

Die regelmäßige Beschäftigung mit einem Problem eines bestimmten Bereichs bewirkt, dass sich das Gehirn in mehreren Phasen an die neuen Anforderungen anpasst. Im Zuge von drei Phasen, der kognitiven, assoziativen und autonomen Phase, wird deklaratives Wissen zu prozeduralem Wissen. In der ersten Phase sind der lernenden Person die Teilziele des Problems und die einzelnen Lösungsschritte bewusst. Der Prozess ist in seinen Einzelschritten im Gedächtnis gespeichert und kann problemlos beschrieben werden. Bei physischen Tätigkeiten verlaufen die Bewegungen in dieser Phase meist unökonomisch und unkoordiniert ab. In der assoziativen Phase wird die Tätigkeit immer schneller ausgeführt und zeichnet sich auch durch eine bessere Koordination und Ökonomie aus. Das Gehirn passt sich weiterhin an, einzelne Schritte werden zu Sequenzen zusammengeschlossen und das Arbeitsgedächtnis dabei entlastet. Die Tätigkeit wird in dieser Phase noch kontrolliert ausgeübt, dennoch sinkt bereits die Reizschwelle. In der letzten Phase tritt die weitgehende Automatisierung ein, wobei der benötigte kognitive Aufwand sinkt. Die Tätigkeit wird deutlich schneller ausgeführt, da die benötigten Lösungsschritte im Gedächtnis bereits gespeichert sind. Vor allem bei monotonen Arbeiten kann die benötigte Konzentration bis auf null sinken. Bei zahlreichen Tätigkeiten kommt es zu einem Grad der Reduzierung, die es erlaubt, bei der Ausführung gleichzeitig andere Aktivitäten auszuüben. In dieser Phase liegt das Wissen der Personen als prozedurales Wissen vor, wobei keine Strategien mehr für das Lösen des Problems angewendet werden müssen. An diesem Punkt des Lernprozesses ist es oftmals schwierig, die einzelnen Schritte in Worte zu fassen, da die benötigten Sequenzen beinahe unbewusst abgerufen werden. Dafür müsste wieder eine Umwandlung des prozeduralen Wissens in die anfangs erwähnte Form des deklarativen Wissens erfolgen (Winter 2011:207ff.).

So stellt sich weiters die Frage, inwiefern die individuelle Begabung eines Menschen eine Rolle spielt. Unter Begabung wird die Fähigkeit eines Menschen verstanden, „die Struktur des Wissens, die verwendeten Strategien und die metakognitiven Kompetenzen je nach Bedarf und Situation anzupassen“ (Chabasse 2009:46). Die Begabungsforschung interessiert sich sowohl für die potenzielle Leistungsfähigkeit einer Person im Allgemeinen also auch auf einen bestimmten Fachbereich bezogen (Winter 2011:210). Somit handelt es sich bei Begabung um eine latente Größe, um das Potenzial bzw. die Kompetenz einer Person (Chabasse 2009:47).

Dieser Faktor ist für die Entwicklung von Expertise bedeutend, denn nicht alle Personen werden rein durch Erfahrung zu Expert_innen. Individuelle Faktoren spielen ebenfalls eine

Rolle (Chabasse 2009:47). Den Erkenntnissen aus der Genetikforschung sowie der Entwicklungs- und Begabungspsychologie zufolge haben Menschen angeborene Dispositionen, die sich günstig auf das Erlernen von bestimmten Tätigkeiten auswirken. Auch Untersuchungen zum Übungsaufwand konnten zeigen, dass es auch bei den gleichen Übungszeiten individuelle Unterschiede in der Leistung gab (Winter 2011:212f.). Vergleicht man die Leistungen von Spitzenexpert_innen, zeichnet sich ebenfalls ab, dass es einen persönlichen maximal erreichbaren Grad an Expertise zu geben scheint (Gruber & Mandl 1992:64).

Dennoch muss bedacht werden, dass das Vorhandensein einer Begabung bzw. Kompetenz noch keine Garantie für die Entwicklung einer Expertise in einem bestimmten Bereich ist (Chabasse 2009:47). So konnten prospektive und retrospektive Studien zu außergewöhnlichen Leistungen aufzeigen, dass zahlreiche nichtkognitive Faktoren eine Rolle für den späteren beruflichen Erfolg bzw. die Leistungsentwicklung spielen. Dazu zählen beispielsweise Aspekte wie soziale Schicht, Konzentration, Ausdauer, Anstrengungsbereitschaft und Unterstützung im familiären und schulischen Bereich (Schneider 1992:106ff.). So können Mängel in der Begabung durchaus kompensiert werden. Faktoren wie Motivation und Übung spielen hierbei ebenfalls eine entscheidende Rolle (Chabasse 2009:47ff.). Denn „when highly experienced individuals are appropriately motivated they are able to improve their objective performance, sometimes dramatically“ (Ericsson 2000:191). Andererseits kann das Vorhandensein einer Begabung wiederum positive Auswirkungen auf die Motivation haben und das Übungsverhalten fördern und zu einer längeren und intensiveren Auseinandersetzung mit einer Tätigkeit führen (Winter 2011:213). Gruber & Mandl (1992) weisen auch auf die mögliche Existenz einer dispositionsabhängigen Motivation hin (Gruber & Mandl 1992:65). So geht eine Begabung möglicherweise mit einer höheren Übungsbereitschaft einher, was wiederum zu einer Leistungssteigerung führt (Winter 2011:213). Idealerweise könnte sich auf diese Weise eine „Aufwärtsspirale aus Motivation - Üben - Erfolg - Motivation - noch mehr Motivation usw. ergeben“ (Winter 2011:213).

Gruber & Mandl (1992:64) sprechen von einer Verknüpfung des Begabungs- und Übungsansatzes und von Leistungsasymptoten, die sich für jede Person für jeden Bereich individuell ergibt. Während diese leicht erreicht werden kann, bedarf es für eine weitere Steigerung viel Training und Einsatz sowie den Erwerb eines immensen Wissens (Gruber & Mandl 1992:64). Je länger sich eine Person jedoch mit einem Gebiet befasst, desto weniger fallen Begabungsunterschiede ins Gewicht (Gruber & Mandl 1992:68).

Dennoch herrscht bei der Frage nach der Bedeutung von Begabung für die Expertise Uneinigkeit. Ericsson (2000:190ff.) beispielsweise kritisiert einige verbreitete Sichtweisen auf Expertise, darunter die angenommenen individuellen Grenzen. Dies stimme nicht überein mit der „observed ability of individuals to improve their performance through training“ (Ericsson 2000:191). So seien Größe und Gewicht die einzigen genetisch bedingten Faktoren, die in diesem Kontext eine Rolle spielen. Ansonsten gäbe es nicht ausreichend Belege für individuelle Eigenschaften, die für die Entwicklung eines Expert_innen-Status erforderlich wären, jedoch nicht durch Üben verbessert werden könnten (Ericsson 2000:192). Ericsson (2000) spricht in diesem Zusammenhang von *deliberate practice* als essenziellen Faktor zur Erlangung von Expertise. In der Expertiseforschung gilt das Modell von Ericsson (2000) sowohl als das einflussreichste als auch umstrittenste Expertisemodell (Winter 2011:210). Im Folgenden soll kurz auf die Prinzipien des Modells und dessen Kritik eingegangen werden, gefolgt vom Schwellenmodell nach Schneider (1992), das Begabung integriert und somit eine angepasste Version des Modells nach Ericsson (2000) darstellt und dessen Kritikpunkte berücksichtigt (Winter 2011:214).

3.1.2. Expertisemodelle nach Ericsson (2000) und Schneider (1992)

Ericsson (2000) beschreibt Erkenntnisse zur *expert performance* aus den Bereichen des Schachspielens, der Musik, der Medizin und des Sports und betont, dass diese in erster Linie „through the engagement in designed training activities, namely deliberate practice“ (Ericsson 2000:193) erreicht wird. Wie bereits erwähnt, lehnt er individuelle Faktoren als Begrenzung in Hinblick auf das Erreichen eines Expert_innen-Status ab (Ericsson 2000:191f.). Die Entwicklung erfolgt schrittweise und kontinuierlich, und auch die talentiertesten Personen brauchen etwa 10 Jahre, um ein hohes Niveau zu erreichen. Dabei ist es entscheidend, klar definierte Aufgaben zum Üben heranzuziehen, Feedback zu erhalten, und die Tätigkeiten zu wiederholen, wobei Übungseinheiten optimalerweise nicht länger als eine Stunde dauern (Ericsson 2000:194f.).

Nach einer gewissen Zeit an Übung und Erfahrung können Menschen grobe Fehler reduzieren und ein akzeptables Niveau erreichen. Die anfangs erforderliche Konzentration sinkt und es kommt zu einer Automatisierung der Tätigkeit. Echte Expert_innen sind jedoch in der Lage, durch gezieltes Üben ihre Leistung stets zu verbessern und dieser Automatisierung

entgegenzuwirken (Ericsson 2000:197f.). Die *expert performance* verbessert sich stets weiter „as a function of more experience coupled with deliberate practice“ (Ericsson 2000:198).

Die Kritik an dem beschriebenen Modell beinhaltet zum einen den Vorwurf, wissenschaftliche Erkenntnisse in Hinblick auf Faktoren, die nicht im Rahmen der *deliberate practice* liegen, zu ignorieren (Winter 2011:212). Das Modell basiere nämlich auf der Annahme, dass lediglich die Intensität und Dauer des Übens ausschlaggebend dafür ist, ob Personen auf ihrem Gebiet zu Expert_innen werden (Winter 2011:211). Zum anderen wird auch auf Schwachpunkte in Hinblick auf die Durchführungs- und Auswertungsmethoden hingewiesen (Winter 2011:212). Denn das Modell basiere auf der Ableitung eines Kausalzusammenhangs, der rein aufgrund der gefundenen Korrelation von Übungsaufwand und der erbrachten Leistung basiert (Winter 2011:213).

Auch Schneider (1992) geht davon aus, dass nichtkognitive Merkmale eine wichtige Rolle spielen, schließt daraus jedoch nicht, dass individuelle Fähigkeiten unbedeutend wären. Denn auch wenn eine Begabung eine außergewöhnliche Entwicklung allein nicht erklären kann, ist auch der Übungsansatz nicht gänzlich als Erklärung ausreichend. So schlägt er ein Schwellenwert-Modell vor, das sowohl basale Fähigkeiten als auch nichtkognitive Aspekte vereint. Sobald die Fähigkeiten einen gewissen Schwellenwert, sprich eine auf überdurchschnittlichem Niveau angesetzte Grenze, überschreiten, entscheiden nichtkognitive Faktoren und die Lernumwelt, ob eine Person Spitzenleistungen erbringt. Damit das möglich wird, sind zwar überdurchschnittliche individuelle Fähigkeiten erforderlich, müssen jedoch nicht im Bereich der Hochbegabung liegen. Wie hoch der Schwellenwert liegt, ist vom Leistungsbereich abhängig und liegt bei komplexeren Bereichen höher (Schneider 1992:119).

3.2. Expertise im Bereich des Dolmetschens

Wie oben dargelegt, scheint auch die Begabung einer Person für das Erlangen von Expertise eine Rolle zu spielen. So zeigt ein Blick in die dolmetschwissenschaftliche Literatur zum Thema Expertise im Dolmetschen, dass sich die Autor_innen stets auch mit der Frage nach Eignungskriterien bzw. Voraussetzungen, die eine Person für den Beruf des Dolmetschens mitbringt, auseinandersetzen (z.B. Chabasse 2009). Moser-Mercer (2000) fasst beispielsweise eine Reihe von Experimenten zusammen, deren Ergebnisse als Hilfe für die Auswahl von Dolmetschkandidat_innen herangezogen werden können. Chabasse (2009) beschäftigt sich in

ihrer Dissertation mit dem Thema Eignungsprüfungen für Studiengänge zum Dolmetschen und setzt sich mit der Frage auseinander, wie ein geeigneter Dolmetscheignungstest erstellt werden kann. Setton (2007) spricht im Zusammenhang von Expertise jedoch von „confusion in interpreting research“ (Setton 2007:63), die möglicherweise zum Teil auf das Übernehmen von Konzepten anderer Disziplinen zurückzuführen ist (Setton 2007:63). Er betont, dass oftmals nicht ausreichend zwischen den unterschiedlichen Fähigkeiten differenziert wird. Grundlegende, allgemeine kognitive Fähigkeiten eines Menschen beinhalten demnach beispielsweise Aspekte wie das Kurzzeitgedächtnis. Zur zweiten Gruppe zählen erlernte Fähigkeiten bzw. Tools, wie das Allgemeinwissen oder Sprachkenntnisse. Schließlich umfasst die aufgabenspezifische Expertise Aspekte wie das Anpassen an den Sprechrhythmus und die Informationsdichte der Ausgangsrede, die Entwicklung einer sprachlichen Gewandtheit oder die Fähigkeit, den für die Kommunikation relevanten Sinn zu erfassen (Setton 2007:64). So kommt Chabasse (2009) auch zu dem Schluss, dass der Begriff „Eignung“ für das Simultandolmetschen besser geeignet ist als jener der Begabung. Ersterer mache es möglich, auch Einflüsse der Umwelt einzubeziehen und die Rolle des Übens zu berücksichtigen (Chabasse 2009:203). Außerdem gäbe es „nicht EINE Begabung für das Simultandolmetschen, sondern mehrere Faktoren, die - je nach Ausprägung und Kombination - den Erfolg bei einer Dolmetscherausbildung mehr oder weniger begünstigen“ (Chabasse 2009:203).

Für die Entwicklung einer Expertise im Dolmetschen sind eine Reihe von Teilfertigkeiten sowie Teilprozessen der Sprachverarbeitung notwendig. Dazu zählen zunächst die Prozesse des Verstehens, wie semantische und syntaktische Verarbeitung des Gehörten. Weiters müssen zwei Sprachsysteme gleichzeitig aktiviert werden. Hinzu kommen die Sprachproduktion, die semantische und syntaktische Verarbeitung sowie phonologische und grammatikalische Kodierung. All diese Prozesse laufen mehr oder weniger simultan ab (Casado et al. 2000:107). Auch Setton (2007:63) betont, dass beim Dolmetschen die Überwachung zweier zeitlich leicht versetzter Reize, sprich der eigenen Stimme und jener der Sprecher_innen, notwendig ist. Mit der Erlangung von Expertise werden diese zwei Tätigkeiten zu einer. Somit handelt es sich nicht mehr um Multitasking, bei dem zwei nicht vereinbare Aktivitäten getrennt ablaufen (Setton 2007:63). Denn Simultandolmetschen „involves monitoring two stimuli [...] which can be collapsed into a single task by forming a single representation (meaning)“ (Setton 2007:63). Expertise im Dolmetschen ist somit „the coordination through practice of normally separate activities (in conversation, we usually alternate speaking and listening)“ (Setton 2007:63).

Da Dolmetschen als ein *ill-defined problem* betrachtet werden kann, das besonderes Wissen erfordert, stellt es für die Expertiseforschung einen herausfordernden Bereich dar (Winter 2011:216). Hinzu kommt die große Bandbreite an möglichen Spezialisierungen und die vergleichsweise kleine Anzahl an Personen, die regelmäßig die gleiche Tätigkeit ausführen (Ericsson 2000:203). Daher werden Erkenntnisse aus anderen Bereichen der Expertiseforschung auf das Dolmetschen übertragen (z.B. Ericsson 2000; Winter 2011). So entwickelt Winter (2011) eine Arbeitshypothese für effizientes Üben zum Expertiseerwerb im Dolmetschen und greift Erkenntnisse aus dem klassischen Instrumentalspiel auf. Trotz einiger Unterschiede, wie beispielsweise der Bedeutung motorischer Abläufe beim Musizieren, oder der Tatsache, dass Musiker_innen ihre theoretischen und praktischen Grundvoraussetzungen zumeist zeitgleich erlangen, während Dolmetscher_innen die Sprachenkenntnisse vor dem Erlernen des Dolmetschens erwerben, gibt es zwischen den beiden Anforderungsprofilen durchaus Parallelen. So handelt es sich in beiden Fällen um eine *ill-defined* und *knowledge-rich* Tätigkeit. Das Resultat beider Tätigkeiten ist individuell und muss zugleich dem Original, sprich der Originalrede bzw. dem Notentext, treu bleiben. Sowohl das Dolmetschen als auch das Musizieren werden jedoch von situationsabhängigen Faktoren beeinflusst, die den Akteur_innen einen gewissen Freiraum einräumen. Weiters zeichnen sich beide Tätigkeiten durch einen kognitiven Aufwand aus, der sich aus verschiedenen, ineinander greifenden Einzelprozessen ergibt. Schließlich müssen sowohl Musiker_innen als auch Dolmetscher_innen stilsicher, flexibel, einführend und stressresistent sein und in schwierigen Situationen Ausweichstrategien heranziehen können. Auch Charakteristika des Berufsalltags wie häufiges Reisen, unterschiedliche Arbeitsbedingungen, das freiberufliche Arbeiten und die damit einhergehende materielle Unsicherheit gelten für beide Tätigkeiten (Winter 2011:218ff.).

Die Forschung an Musiker_innen hat gezeigt, dass besonders fähige Personen in der Lage waren, geschwärzte Takte logisch zu ergänzen, ähnlich dem Antizipieren beim Dolmetschen. Auf diese Weise kann das Arbeitsgedächtnis entlastet werden. Durch die Automatisierung, die bei Musiker_innen besonders am Anfang der Ausbildung eintritt, können sie sich folglich besser auf Feinheiten und Abweichungen konzentrieren und ihre Leistung steigern. Ähnliches wird auch für das Dolmetschen vermutet. So ermöglicht das Verinnerlichen von Kollokationen und Redensarten in weiterer Folge eine Konzentration auf Abweichungen (Winter 2011:221).

Aus den beschriebenen Erkenntnissen leitet Winter (2011) Hypothesen in Hinblick auf sinnvolles Üben zum Erwerb von Dolmetschexpertise ab. Sie schlägt das Einstudieren von häufig verwendeten Kollokationen und Formulierungen in unterschiedlichen Varianten vor.

Weiters sollen sich Dolmetscher_innen mit den Redner_innen, ihrer Biografie und ihrer Person, sowie verfügbaren früheren Redebeiträgen auseinandersetzen. So kann bereits Wissen erworben werden, was später zu einer Entlastung beiträgt und eine Konzentration auf das Lösen von Problemen ermöglicht. Schließlich wird Studierenden empfohlen, Texte in kürzerer Zeit nochmals zu dolmetschen, um fehlerhafte Stellen mit besseren Lösungen zu „überschreiben“ und die Aufmerksamkeit auf unterschiedliche Aspekte zu lenken (Winter 2011:222f.).

Wie bereits erwähnt, kommt in der Dolmetschwissenschaft auch das sogenannte Expert_innen-Noviz_innen-Paradigma zur Anwendung, wonach die Leistungen von Anfänger_innen mit jenen von Expert_innen verglichen werden (Chabasse 2009:53ff.). Dieser Ansatz ist beispielsweise bei Moser-Mercer (2000) oder Sunnari (2003) zu finden. Zunächst kann man sagen, dass professionelle Dolmetscher_innen ein höheres Niveau an Leistung und Fähigkeiten aufweisen als bilinguale Anfänger_innen oder Dolmetscher_innen, die sich noch in der Ausbildung befinden (Sunnari 2003:237). Doch was genau unterscheidet Noviz_innen von Expert_innen? Während Anfänger_innen nur wenig bis keine Erfahrung in einem Bereich haben, konnten Expert_innen aufgrund von jahrelanger Erfahrung ein hohes Leistungsniveau erreichen (Casado et al. 2000:108). Während die Tätigkeit bilingualer Anfänger_innen noch als Multitasking bezeichnet werden kann, da sie zwei normalerweise widersprüchliche Aktivitäten, Sprechen und Hören, gleichzeitig bewältigen müssen, passen Expert_innen ihren Output an das Gehörte an (Setton 2007:64) und fokussieren sich „on a single unfolding representation of meaning, and a single goal: communication“ (Setton 2007:64). Weiters haben sie sich die Fähigkeit des *self-monitoring* angeeignet. Expert_innen sind dazu in der Lage, mehr Ressourcen für die Output-Kontrolle aufzuwenden und sie sind sich der sprachlichen Merkmale ihrer Textproduktion bewusst (Setton 2007:65). Im Vergleich zu Anfänger_innen sind sie auch in der Lage, Strategien einzusetzen, um den Inhalt des Ausgangstextes effektiver zu übertragen. Dabei können sie überflüssige und redundante Informationen herausfiltern und können effizienter dolmetschen. Ihre Dolmetschung ist genauer und „wirtschaftlicher“, mit weniger falschen Satzanfängen und ungeschickten Strukturen. Genauigkeit und Wirtschaftlichkeit der Dolmetschleistung wiederum sind eng mit gängigen Qualitätskriterien verknüpft (Sunnari 2003:244f.).

Die bisherigen Ausführungen zeigen deutlich, dass Dolmetschen „a complex cognitive skill“ (Casado et al. 2000:107) darstellt. Da die Erlangung von Expertise ein Entwicklungsprozess ist (Hoffmann 1997:218), stellt das Üben einen zentralen Faktor dar (Ericsson 2000). Ericsson (2000:189) merkt an, dass Personen nach ihrer Ausbildung beim Eintritt ins Berufsfeld zunächst überfordert und auf die Hilfe anderer angewiesen sind. Erst

nach Monaten oder Jahren erlangen sie „a level of acceptable proficiency and are able to work independently and further develop their professional reputation” (Ericsson 2000:189). Doch auch wenn zusätzliche Jahre der Berufspraxis erforderlich sind, um vollwertige Expert_innen zu werden, gelten Studierende nach ihrem Abschluss bereits gewissermaßen als Expert_innen (Casado et al. 2000:108). So sei im Bereich des Dolmetschens anerkannt, „that the student obtaining his final diploma can call himself an expert with some degree of justification” (Casado et al. 2000:108). Die Dolmetschdidaktik hat nämlich zum Ziel, voll einsatzfähige Dolmetscher_innen auszubilden, die nach ihrem Abschluss unmittelbar dazu befähigt sind, am ersten Arbeitstag professionelle Dolmetschleistungen zu erbringen (Déjean Le Féal 1998:363).

Mit welchen Schwierigkeiten die Qualitätsbewertung im Konferenzdolmetschen einhergeht, wurde bereits in Kapitel 2.3. erläutert. Nun stellt sich die Frage, welche Rolle die Expertise bzw. ihre Erforschung hierbei eine Rolle spielt. Sunnari (2003:244) betont, dass wenn wir mehr Einsicht in die Fähigkeiten und Fertigkeiten im Dolmetschen gewinnen, „about how they develop and change in the course of practical work and training, it might be possible to move away from such ‘shopping-list’ type attributions to more specific criteria reflecting the expected or required level of skill” (Sunnari 2003:244). Denn im didaktischen Bereich bewerten Lehrende ihre Studierenden oftmals aus dem Blickwinkel ihrer eigenen Praxis, was einer Bewertung von Noviz_innen anhand einer Expert_innen-Skala gleicht (Moser-Mercer 2000:339).

Betrachtet man Expertise hingegen in Hinblick auf ihre Rolle beim Urteilen, führt dies beispielsweise zur FAIT nach Schwarz (2011). Laut dieser Theorie nimmt der Einfluss von Stimmungen, der zentrales Forschungsinteresse dieser Arbeit ist, mit zunehmendem Expertisegrad ab (Schwarz 2011:297). Das Einbeziehen der FAIT nach Schwarz (2011) führt in den Bereich der Affekte, auf den im nächsten Kapitel eingegangen werden soll.

4. Affekte

In der Forschung wird Affekten und ihrer Rolle für unser Leben immer mehr Aufmerksamkeit zuteil (Koskinen 2020:3). Lange Zeit wurde der Einfluss von Emotionen auf die Kognition in der Psychologie jedoch ignoriert und man war bemüht sicherzustellen, dass sich untersuchte Personen in einem neutralen Gefühlszustand befanden (Eysenck & Keane 2020:715). Da in den 1950ern der kognitive Ansatz in der Psychologie vorherrschte, der den Fokus auf formale Strukturen lenkte, die das Gedächtnis sowie Urteilen und Denken steuern, wurden Affekte und ihre Bedeutung im Allgemeinen somit außer Acht gelassen. Forschung zu diesem Thema wurde eher in Subdisziplinen durchgeführt, wobei die verschiedenen Ansätze nicht miteinander verknüpft wurden. Mit der Erkenntnis, dass die Theorien der Kognitionspsychologie auch Schwächen und Grenzen aufweisen, wurde nun versucht, den Blick auf den Affektbereich zu erweitern und andere Phänomene miteinzubeziehen. Doch auch der gesellschaftliche Kontext ab Ende der 1960er mit dem Fokus auf Aspekte wie die individuelle Entfaltung und die persönliche Erfahrung, sowie die allgemeine Ablehnung des Materialismus begünstigten das Erwachen des Interesses am Bereich der Affekte (Briner et al. 2000:22ff.).

Der Affektbegriff umfasst ein breites Spektrum an Phänomenen und beinhaltet auch Stimmungen als Untergruppe (Briner et al. 2000:15f.). Um den Stimmungseinfluss auf die Evaluation von Dolmetschleistungen betrachten zu können, ist also ein interdisziplinärer Blick notwendig. Dieser führt in den Bereich der Sozialpsychologie und konkreter in die Stimmungsforschung. Dabei steht der Einfluss von Affekten auf die kognitiven Verarbeitungsprozesse im Fokus (Behr 2013a:219). So soll im Folgenden zunächst auf den Affektbegriff sowie den Begriff der Stimmung eingegangen werden. Dem soll ein Blick in übersetzungs- bzw. dolmetschwissenschaftliche Arbeiten zur Auseinandersetzung mit Affekten folgen. Schließlich soll die Bedeutung von Stimmungen auf das Urteilen betont und der Bogen zum empirischen Versuch dieser Arbeit gespannt werden, indem auf die *Feelings-as-information theory* nach Norbert Schwarz (2011), sowie die Studie von Martina Behr (2013a) genauer eingegangen wird.

4.1. Definitionen und Abgrenzungen

Aus psychologischer Sicht können mentale Funktionen auf grundlegender Ebene allgemein gesprochen in drei Kategorien eingeteilt werden. Die erste beinhaltet das Denken bzw. die Kognition. Die zweite Grundfähigkeit ist jene der Konation und meint Wünsche und das

Wollen. Die letzte Kategorie umfasst Affekte, sprich das Fühlen, und schließt somit auch Stimmungen mit ein. Auch wenn diese hauptsächlich zur Grundfähigkeit der Affekte gezählt werden können, überschneiden sie sich auch mit anderen der drei Grundkategorien, da sich beispielsweise Stimmungen und das Denken gegenseitig beeinflussen. Zusätzlich wirken sich Stimmungen auch auf Handeln und Motivation von Menschen aus (Briner et al. 2000:15). Allgemein formuliert meint der Begriff Affekt „mentale Zustände, die *evaluative Gefühle* umfassen, mit anderen Worten [...] psychische Bedingungen, unter denen sich die Person wohl oder unwohl fühlt und entweder begrüßt oder ablehnt, was geschieht“ (Briner et al. 2000:15). So kann der Begriff Affekt in der Psychologie als Oberbegriff angesehen werden, der Emotionen und Stimmungen umfasst (Briner et al. 2000:15f.; Eysenck & Keane ⁸2020:716) und sowohl Phänomene wie allgemeine Langeweile oder ein vages Unbehagen als auch ausgeprägte Wut miteinbezieht (Briner et al. 2000:15f.). Positive Affekte umfassen somit sowohl positive Stimmungen als auch positive Emotionen, während negative Affekte die negativen Ausprägungen dieser Zustände erfassen (Eysenck & Keane ⁸2020:716).

Ein Blick in die Fachliteratur macht deutlich, dass eindeutige Definitionen bzw. Abgrenzungen im Bereich der Affekte jedoch oftmals schwer möglich sind. Erklärungsansätze bzw. Definitionsversuche in diesem Kontext werden oftmals in unterschiedlichen Forschungsbereichen bzw. Disziplinen vorgenommen oder sind an Schnittstellen anzusiedeln. So überschneiden sich die unterschiedlichen Begriffe in ihren Inhalten und greifen ineinander und können nicht immer klar getrennt werden (Behr 2013a:221). Mit anderen Worten: „In the conceptual jungle surrounding affects and emotions, the same term can be used to mean different things, and confusing synonyms abound.“ (Koskinen 2020:12) Auch die Subdisziplinen der Psychologie, die sich abseits des vom kognitiven Ansatz geprägten Mainstreams der Psychologie mit der Thematik beschäftigten, entwickelten jeweils ihre eigenen Erklärungen zu Affekten. Es wurde jedoch kein ernsthafter Versuch unternommen, die unterschiedlichen Erkenntnisse miteinander in Verbindung zu bringen (Briner et al. 2000:22f.). Doch nicht nur in der Psychologie, sondern auch in verschiedenen anderen wissenschaftlichen Disziplinen wurde der Affektbegriff auf unterschiedliche Art und Weise gebraucht (Wassmann 2017:166).

An dieser Stelle kommt hinzu, dass es dem interdisziplinären Forschungsfeld im Bereich der Affekte an einer historischen Perspektive hinsichtlich der Verwendung und Entwicklung der verschiedenen Begriffe in den unterschiedlichen Sprachen fehlt. Wissenschaftliche Arbeiten werden in der heutigen Zeit größtenteils auf Englisch publiziert, wobei es zu synonymen Verwendungen von Begriffen kommt, die in den jeweiligen Sprachen

semantisch nicht deckungsgleich sind. So haben sich unterschiedliche Begrifflichkeiten wie Gefühl, Empfindung oder auch Affekt in den drei Sprachen Englisch, Französisch und Deutsch unterschiedlich entwickelt und wurden zu unterschiedlichen Zeitpunkten in der Geschichte jeweils auch anders gebraucht. Dieser Aspekt wurde beim gegenseitigen Rezipieren wissenschaftlicher Arbeiten, die in anderen Sprachen verfasst wurden, oft nicht bedacht, weshalb es auch zu Fehlern in der Übersetzung kam. Unterschiede, die beispielsweise im Deutschen wichtig waren, gingen im Zuge des Übersetzens ins Englische verloren. Weiters wurden und werden Aspekte bzw. Fragestellungen, die nicht auf Englisch publiziert wurden, teilweise ignoriert (Wassmann 2017). Schließlich werden Begriffe auch je nach Präferenz verwendet. So wählt beispielsweise Koskinen (2020) gezielt den Begriff Affekt, mit dem Ziel, einen ungewöhnlichen Ausdruck zu verwenden und bekannte Bedeutungen und Interpretationen zu überwinden (Koskinen 2020:1).

Selbst wenn für die vorliegende Arbeit die psychologische Sichtweise auf Stimmungen zentral ist, sei an dieser Stelle hervorgehoben, dass andere Bereiche ebenso an Affekten interessiert sind. So schenken auch die Kultur- und Sozialwissenschaften sowie die Soziologie dem Thema der Affekte und deren sozialer Beschaffenheit mehr Aufmerksamkeit (Koskinen 2020:18). Doch auch manche Sozialpsycholog_innen betrachten beispielsweise Emotionen als soziale Konstruktionen (Scherer ⁴2002:203), die „ausschließlich eine kulturell bestimmte oder sozial konstruierte Wirklichkeit besitzen“ (Scherer ⁴2002:203). Doch unabhängig davon, wie Affekt definiert wird,

as a nonconscious visceral reaction or a social force or something in between – it is part of our human condition: our systems of observing and responding to our surroundings are not only rational but also affective, and our affective responses allow us to interpret, assess and judge the developments we witness and take part in. (Koskinen 2020:179)

Um speziell den Begriff der Stimmung zu definieren, wird in der Psychologie eine Abgrenzung bzw. Differenzierung zu Emotionen vorgenommen (Briner et al. 2000:16-20; Eysenck & Keane ⁸2020:716; Scherer ⁴2002:171). Da die vorliegende Arbeit speziell an Stimmungen und deren Einfluss interessiert ist, soll daher nun auf den Stimmungsbegriff und seine Unterschiede zum Begriff der Emotion eingegangen werden.

4.1.1. Stimmungen vs. Emotionen

Der Begriff „Stimmung“ ist uns allen geläufig und wird im Alltag selbstverständlich verwendet. Unsere Idee bzw. Definition der Stimmungen basiert auf den individuellen Erfahrungen, die wir mit ihnen machen. Wir sprechen oft davon, nicht in der richtigen Stimmung zu sein oder davon, dass wir eine bestimmte Laune haben. Bereits der Alltagsbegriff spricht zwei zentrale Aspekte des mentalen Zustandes der Stimmung an. Zum einen wird von angenehmen sowie unangenehmen Ausprägungen ausgegangen. Zum anderen begünstigen sie, dass wir bestimmte Handlungen vollziehen oder diese unterlassen. So werden Stimmungen im alltäglichen Gebrauch oft nach ihrer qualitativen Ausprägung sowie nach ihrer Intensität beschrieben. Dennoch werden die im Alltag verwendeten Stimmungsbegriffe nicht genau eingegrenzt, greifen daher verschiedene Phänomene auf und basieren auch manchmal auf Mythen (Briner et al. 2000:11-14).

Obwohl die im Alltag geläufige Stimmungsdefinition weit gefasst ist und nicht präzise eingegrenzt ist, leitet die Psychologie ihre Definition vom alltäglichen Verständnis der Stimmung ab. Dennoch wird versucht, das Phänomen durch ein klar eingegrenztes Konzept zu beschreiben. Aus wissenschaftlicher Sicht kann Stimmung als Unterkategorie des bereits aufgegriffenen Begriffs des Affektes gesehen werden. Eine mögliche Definition bzw. Erfassung erfolgt durch die Abgrenzung und den Vergleich zum Begriff der Emotionen (Briner et al. 2000:15f.). Einfach gesagt kann Emotion als „a short-lived affective state typically triggered by a specific event“ (Eysenck & Keane ⁸2020:807) angesehen werden. Hiervon ausgehend können Stimmungen als „[a] state resembling emotion but generally longer lasting, less intense and of unknown cause“ (Eysenck & Keane ⁸2020:813) definiert werden. Unterschiedliche Theorien in der Psychologie haben verschiedene Stimmungs- bzw. Emotionsmerkmale hervorgehoben. So können diese zwei Konzepte allgemein anhand von sechs Merkmalen unterschieden werden. Dazu gehören die Dauer, das zeitliche Muster, die Intensität, die Verursachung, die Signalfunktion und die Gerichtetheit der beiden Phänomene (Briner et al. 2000:16-20).

Zunächst kann festgehalten werden, dass Stimmungen im Vergleich zu Emotionen über eine längere Zeit anhalten (Briner et al. 2000:16,19; Eysenck & Keane ⁸2020:716; Scherer ⁴2002:171). Auch wenn es durchaus Ausnahmen zu geben scheint, so sind Stimmungen von längerer Dauer, wohingegen Emotionen kürzer bestehen bleiben. Während letztere oftmals bereits nach Minuten oder gar Sekunden vorbei sind, können Stimmungen über einen Zeitraum von Stunden oder Tagen, bis hin zu Wochen anhalten. Auch wenn Emotionen über eine längere

Periode anzuhalten scheinen, wie beispielsweise Trauer, so ist eine mögliche Erklärung hierfür, dass diese Emotion immer wieder abnimmt und erneut zunimmt, so dass es tatsächlich zu mehreren wiederholten emotionalen Reaktionen kommt (Briner et al. 2000:16).

Weiters können die zwei Phänomene anhand ihrer typischen zeitlichen Muster unterschieden werden. Die längerfristig andauernden Stimmungen setzen graduell ein, verändern sich auch langsamer und können als kontinuierlich bezeichnet werden. Emotionen hingegen scheinen einen eindeutigen Anfangspunkt zu haben, setzen schnell ein und klingen auch schneller wieder ab. Sie können daher als akut und phasisch bzw. episodisch bezeichnet werden (Briner et al. 2000:16f.,19).

Ein weiteres mögliches Unterscheidungsmerkmal stellt die Intensität dar. Emotionen zeichnen sich im Vergleich zu Stimmungen durch einen höheren Intensitätsgrad aus, während letztere allgemein als weniger intensiv charakterisiert werden (Briner et al. 2000:17; Eysenck & Keane ⁸2020:716; Scherer ⁴2002:171). Folglich „sind Emotionen offenbar relativ stark, brennen aber schneller aus, während Stimmungen schwächer und von eher nagender Beharrlichkeit sind“ (Briner et al. 2000:17). Emotionen gelten vergleichsweise als verzehrender, wobei auch bei diesem Unterscheidungskriterium Ausnahmen bestehen (Briner et al. 2000:17). Da sie intensiver als Stimmungen sind, erregen Emotionen aber auch eher unsere Aufmerksamkeit (Eysenck & Keane ⁸2020:716).

Stimmungen können von Emotionen jedoch auch danach unterschieden werden, welche Ursachen ihnen zugrunde liegen bzw. wodurch sie ausgelöst werden. Während Emotionen meist durch einen konkreten Auslöser hervorgerufen werden, sind Stimmungen zumeist auf unklare Ursachen zurückzuführen (Eysenck & Keane ⁸2020:716; Scherer ⁴2002:171). So sind es spezifische Ereignisse, die zeitlich festgelegt sind, die Emotionen zugrunde liegen (Briner et al. 2000:18), wohingegen eine Stimmung „durch eine Verkettung kleinerer Ereignisse, durch anhaltende äußere Bedingungen und/oder durch innere metabolische oder kognitive Prozesse“ (Briner et al. 2000:18) zustande kommt. Einer ähnlichen Sichtweise zufolge bezieht sich eine Emotion auf einen spezifischen Inhalt oder ein konkretes Ereignis, das persönlich erfahren wird, wohingegen Stimmungen von allgemeinen, individuellen Umständen im Leben einer Person abhängig sind. So kann der Verlust einer nahestehenden Person Trauer auslösen, während depressive Stimmungen auf Unzufriedenheit allgemein zurückgeführt werden können (Briner et al. 2000:18).

Auch die Signalfunktion von Emotionen und Stimmungen bzw. die Art der Informationen, die diese zwei Phänomene liefern, werden als ein mögliches Unterscheidungskriterium angesehen. Demnach informieren Emotionen uns über eine Situation

bzw. deren Zustand, sprich über die Außenwelt. Stimmungen dagegen geben uns Signale bezüglich unseres inneren Zustandes, sprich über uns selbst (Briner et al. 2000:18). Durch eine Emotion erhalten wir demnach die Information, was „in der äußeren Situation richtig oder falsch läuft, während eine Stimmung uns mitteilt, ob wir über genügend *innere* Kräfte verfügen, um eine bestehende Herausforderung zu meistern oder nicht“ (Briner et al. 2000:18). Durch die erlebten Emotionen und Stimmungen erhält das Bewusstsein somit Informationen über die Beziehung, die man zu der Welt hat. Dennoch ist es schwierig, dieses Merkmal auf alle Situationen anzuwenden, da es nicht immer einfach ist, zwischen Ereignissen, die sich auf das Selbst beziehen, und Ereignissen, die situationsbezogen sind, zu differenzieren (Briner et al. 2000:18).

Schließlich gelten Stimmungen als diffus und ungerichtet (Briner et al. 2000:19; Scherer 2002:171) während Emotionen sich auf etwas konkretes richten (Briner et al. 2000:19). Briner et al. (2000:19) merken an, dass die oben erläuterte Unterscheidung nach der Verursachung zu vereinfacht ist, und präziser durch das Merkmal „Gerichtetheit“ erfolgen kann. Bei diesem Unterscheidungsfaktor stehen die Richtung bzw. der Inhalt von Stimmungen und Emotionen im Zentrum, weniger die Ursache. Somit ist in diesem Kontext die Unterscheidung zwischen dem Objekt einer Emotion und ihrer Ursache zentral. Nach dieser Auffassung richten sich Emotionen auf ein bestimmtes Objekt, werden aber nicht zwingend durch ein spezifisches Ereignis verursacht bzw. bedingt. Denn es ist ein Zusammenspiel von mehreren situativen und individuellen Aspekten, das zur Verursachung von Emotionen führt. Stattdessen kann von einer Gerichtetheit gesprochen werden, wobei sich dieses Objekt, auf das die Emotion gerichtet ist, mit den zu Grunde liegenden Ursachen zwar überlappen kann, aber in seltenen Fällen ident ist. Während Emotionen sich durch einen spezifischen Inhalt auszeichnen, sind Stimmungen auf nichts Konkretes gerichtet. Eine Stimmung spiegelt eher das allgemeine, diffuse Befinden bzw. die generelle positive oder negative Gemütsverfassung wider. Somit können Stimmungen in den unterschiedlichsten Situationen beeinflussen, wie wir reagieren. Eine Stimmung stellt also einen unfokussierten Wertzustand dar, der sich auf unterschiedlichste Objekte beziehen kann, wohingegen Emotionen fokussierte Wertzustände darstellen. Mit diesem Unterscheidungsansatz hängt auch die Handlungstendenz zusammen. Emotionen richten sich auf ein Objekt und umfassen oftmals eine gewisse Tendenz zum Handeln, die sich auf das gerichtete Objekt bezieht. Dies ist bei Stimmungen nicht der Fall. Sie können zwar eine gewisse Neigung zu einem Verhalten oder eine Abneigung gegen eine Handlung mit sich bringen, führen aber zu keiner konkreten Handlungsweise (Briner et al. 2000:19f.).

Diese letzte mögliche Abgrenzung bietet laut Briner et al. (2000) auch eine mögliche Erklärung für die anderen, deskriptiven Unterscheidungsmerkmale wie zum Beispiel Dauer und Intensität. Denn wenn Emotionen sich durch die Gerichtetheit auf ein spezifisches Objekt auszeichnen, hängt es auch von der Anwesenheit dieses Objekts ab, wie lange die Emotion andauert. Ist das Objekt nicht mehr präsent oder gewöhnt man sich daran, kommt es zu einer Auflösung der Emotion. In diesem Kontext ist auch der Druck anders, sich mit dem Objekt auseinander zu setzen, was einen Einfluss auf die erlebte Intensität hat. Bei Stimmungen kann man hingegen von einer Art Eigenleben sprechen. Sie nehmen zu und ab, abhängig von der generellen individuellen Befindlichkeit und richten sich eben nicht auf einen konkreten Lebensaspekt (Briner et al. 2000:20).

Dennoch sei nochmals hervorgehoben, dass es trotz der Abgrenzungsmöglichkeiten keine scharfe Unterscheidung zwischen Emotionen und Stimmungen gibt (Briner et al. 2000:20f.; Eysenck & Keane ⁸2020:716). Sowohl Emotionen als auch Stimmungen umfassen angenehme sowie unangenehme Affektzustände und weisen folglich auch Gemeinsamkeiten auf und auch ihre Steuerung erfolgt teils durch ähnliche Prozesse. Manchmal werden auch dieselben Adjektive verwendet, um Stimmungen und Emotionen zu beschreiben. Schließlich können gewisse Stimmungszustände aus bestimmten Emotionserfahrungen folgen. Umgekehrt können kleinere emotionale Erlebnisse zusammenkommen und zu einer länger andauernden Stimmung führen. Somit bleiben Emotionen und Stimmungen prinzipiell verwandte Konzepte (Briner et al. 2000:20f.). So beziehen sich auch die unterschiedlichen Darstellungen der Affektstruktur sowohl auf Stimmungen als auch auf Emotionen (Briner et al. 2000:33-45). Auf die Struktur der Affekte soll im Folgenden näher eingegangen werden.

4.1.2. Affektstruktur

Die unterschiedlichen Affektzustände bzw. die Struktur von Affekten wird durch verschiedene Modelle von Gefühlsdimensionen dargestellt. Diese beziehen sich meistens auf das Gefühlserleben allgemein und gelten sowohl für Emotionen als auch Stimmungen. Sie basieren auf einer langen Forschungstradition, in der verschiedene Methoden angewandt wurden, um die Struktur der Affekte abzubilden. Zu diesen Techniken zählen die Analyse von Selbstbeobachtungen, Studien zu Gesichtsausdrücken und Wortbedeutungen, sowie Selbstbeurteilungen über Erfahrungen im Alltag (Briner et al. 2000:32-43).

Die Arbeit von Wundt (1897) war eine der ersten, die Affekte anhand von Dimensionen beschrieb. Sie zählt zu den introspektiven Studien und basiert auf seinen subjektiven

Erfahrungen. Die Erkenntnisse brachten drei Gefühlsdimensionen hervor: Lust/Unlust, Spannung/Lösung und Erregung/Beruhigung. Obwohl das Modell durchaus Einschränkungen aufweist, stimmten Untersuchungen, die zu späteren Zeitpunkten mit anderen Methoden durchgeführt wurden, mit diesen frühen Erkenntnissen überein (Briner et al. 2000:33f.). Andere Studien basierten auf Analysen von Gesichtsausdrücken und wurden beispielsweise mit Fotos durchgeführt, die es in Kategorien einzuordnen galt. Eine Einschränkung stellt an dieser Stelle jedoch die Tatsache dar, dass einige Gefühle womöglich mit keiner spezifischen Mimik zusammen auftreten und somit von Anfang an nicht einbezogen werden. Außerdem wird auch argumentiert, dass Stimmungen nicht zwingend mit einem spezifischen Gesichtsausdruck zusammenhängen. Untersuchungen zu den verschiedenen Bedeutungen von Emotions- und Stimmungswörtern hingegen könnten die Struktur der Affekte womöglich umfassender darstellen, da die wichtigen Unterschiede zwischen den Gefühlen mit höherer Wahrscheinlichkeit in der Sprache verschlüsselt als auf den Gesichtern abgezeichnet sind. Studien, bei denen Selbstbeurteilungen durchgeführt werden, beziehen sich jedoch auf das aktuelle Erfahren bzw. Erleben von Affekten (Briner et al. 2000:34-39). Dabei handelt es sich um die am häufigsten angewandte Methode der Stimmungsmessung (Briner et al. 2000:53).

Die verschiedenen, soeben angeführten Methoden, deren Erkenntnisse zur Konzipierung von Modellen der Affektstruktur herangezogen wurden, weisen jeweils unterschiedliche Vor- sowie Nachteile auf. Sie führten meist zur Aufstellung von zwei bis drei Grunddimensionen des Affekts, die sich größtenteils überschneiden. Heute herrscht Konsens darüber, dass Affektzustände prinzipiell durch zwei Dimensionen beschrieben werden können, was durch zahlreiche Untersuchungen untermauert werden konnte. Dazu zählt zum einen die Dimension Lust/Unlust, anhand derer Affekte in angenehm und unangenehm eingeteilt werden können. So können auch Stimmungen in gut bzw. angenehm oder schlecht bzw. unangenehm eingeteilt werden. Stimmungen kommt somit eine evaluative Bedeutung zu. Bei der zweiten Dimension handelt es sich um jene der Aktivierung. Sie schreibt den Affekten bzw. Stimmungen unterschiedliche Ausprägungen in Hinblick auf den Aktivitäts- bzw. Erregungsgrad zu. Heute ist man sich einig, dass dies die zwei Dimensionen sind, die der Affektstruktur zugrunde liegen. Darüber hinaus herrscht Uneinigkeit, ob weitere Unterteilungen notwendig sind (Briner et al. 2000:31-39). Auch wenn nicht gesagt werden kann, dass diese zwei Dimensionen alles umfassen, was Affekte ausmacht, so scheinen sie dennoch zentral zu sein und schon länger eine Rolle in der Psychologie zu spielen (Mendelsohn et al. 1989:493f.). Dennoch weist diese Struktur auch Grenzen auf. So können aktuelle Skalen mit zwei Dimensionen beispielsweise nicht zwischen den zwei besonders intensiven Affekten

ängstliche Stimmung und gereizte Stimmung unterscheiden. Für bestimmte Untersuchungszwecke scheinen daher wohl andere, gezieltere Herangehensweisen sinnvoll (Briner et al. 2000:55). Da jedoch im allgemeinen Einigkeit über diese zwei Dimensionen herrscht, und die vorliegende Arbeit lediglich zwischen positiver und negativer Stimmung unterscheiden will, erscheint es sinnvoll und ausreichend, sich auf die zwei grundlegenden Dimensionen Lust/Unlust und Aktivierung zu beziehen.

Ein solches Modell zur Struktur der Affekte wurde von Russell (1980) entwickelt. Sein Circumplex-Modell basiert auf den zwei Dimensionen der Aktivierung und des Wohlbefindens (Lust/Unlust). Seine These besteht darin, „that affective states are, in fact, best represented as a circle in a two-dimensional bipolar space“ (Russell 1980:1161f.). Russell (1980) geht von einer kognitiven Struktur des Affekts aus, auf die in verschiedene Situation zurückgegriffen wird, darunter beim Erleben von Affekten selbst. Diese Struktur kann in einem Zirkumplex zusammengefasst werden, das aus zwei im rechten Winkel aufeinander stehenden Achsen besteht, die den zwei Gefühlsdimensionen Lust/Unlust und Erregung/Schläfrigkeit entsprechen. Auch wenn diese zwei Dimensionen nicht alles erfassen können, sprechen die Ergebnisse mehrerer Studien und unterschiedlicher Wissenschaftler_innen für die Heranziehung dieser zwei Dimensionen. Jedes Affektwort kann als Kombination der Lust/Unlust- und der Aktivierungsdimension definiert und sinnvoll entlang des Umfangs angeordnet werden. So fallen nicht alle Gefühlswörter auf die Achsen, sondern auch in den durch die rechten Winkel definierten Raum, sprich in die Quadranten. Das Modell besteht insgesamt aus acht Variablen. Vier davon stellen die Enden der zwei Gefühlsdimensionen dar und werden als *arousal/sleepiness* und *pleasure/misery* bezeichnet. Darüber hinaus sind vier weitere Kategorien Teil des Modells, wobei sie keine eigene Dimension darstellen, sondern dabei helfen, den Raum zu definieren. Dabei handelt es sich um die Kategorien *contentment/distress* und *excitement/depression* (Russell 1980:1161-1164). Anders gesagt handelt es sich beim Circumplex um

eine zweidimensionale, kreisförmige Struktur, bei der die Items hohe Korrelationen mit Items aufweisen, die auf der Kreislinie dicht neben ihnen liegen, Null-Korrelationen ($r=0,0$) mit Items, die auf dem Circumplex im 90 Grad-Winkel zu ihnen stehen, und perfekte negative Korrelationen ($r=-1,0$) mit den Items, die ihnen auf der Kreislinie direkt gegenüberliegen. (Briner et al. 2000:40)

Diese angenommene Affektanordnung wurde in verschiedenen Studien mit unterschiedlichen Erhebungs- und Auswertungsmethoden getestet. Dabei wurden sowohl Selbstberichte als auch

semantische Anordnungen von Affektwörtern ausgewertet. Die Untersuchungen bestätigten im Grunde die vorgeschlagene zirkuläre Anordnung und kamen zu ähnlichen Ergebnissen (Russell 1980).

Basierend auf diesem Modell einer zirkulären Affektstruktur entwickelte Russell gemeinsam mit anderen Wissenschaftler_innen das Affektgitter (*Affect Grid*). Dieses stellt ein Messinstrument dar, anhand dessen Affekte auf den beiden Dimensionen Lust/Unlust (*pleasure/displeasure*) und Aktivierung (*arousal/sleepiness*) erhoben werden können. Es entspricht somit im Prinzip den acht Kategorien des Zirkumplexes, wobei ein bisschen andere Begriffe gewählt wurden (Mendelsohn et al. 1989:493ff.). Abbildung 1 zeigt das Affektgitter nach Mendelsohn et al. (1989).

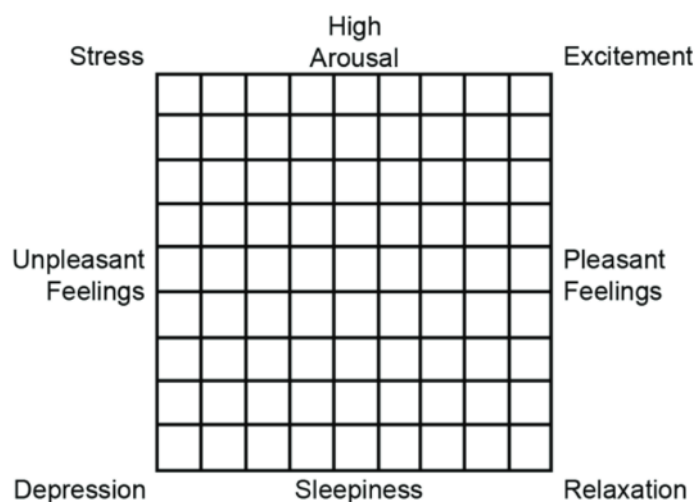


Abbildung 1: Affektgitter nach Mendelsohn et al. (1989:494)

Das Instrument kann zur Bewertung der gegenwärtigen Stimmung, zum Bewerten von einzelnen Gesichtsausdrücken oder einzelnen Affektwörtern verwendet werden. Die Versuchspersonen setzen eine Markierung im Gitter (Mendelsohn et al. 1989:493). Die Autor_innen beschreiben die Methode als „suitable for any study that requires judgments about affect of either a descriptive or a subjective kind“ (Mendelsohn et al. 1989:493).

Basierend auf dem Affektgitter nach Mendelsohn et al. (1989) entwickelte Behr (2013a) ihren Fragebogen, der im Zuge ihrer Erhebung zum Einfluss von Stimmungen bei der Evaluation im Konferenzdolmetschen herangezogen wurde (Behr 2013a:288-291). Dies führt uns somit wieder zurück in den Bereich der Translationswissenschaft. Somit soll nun ein kurzer Einblick zum Interesse der Translationswissenschaft in Hinblick auf Affekte gegeben werden, bevor anschließend die Auswirkungen von Stimmungen näher betrachtet werden.

4.2. Affekte in der Translationswissenschaft

Emotionen sind eng mit Sprache verbunden (Hubscher-Davidson 2018:9). Affekte waren auch zu einem gewissen Grad stets Teil der Übersetzungs- und Dolmetschwissenschaft. Dennoch waren sie selten explizites Forschungsinteresse und haben erst an Bedeutung gewonnen, als andere, benachbarte Forschungsbereiche neue Erkenntnisse zu diesem Thema generiert haben. Ein großer Teil der Forschung wurde im Bereich der kognitiven Übersetzungswissenschaft hervorgebracht und etablierte eine psychologische Sichtweise auf Affekte. Daher geht aktuelle Forschung vor allem von Affekt als etwas im Inneren einer Person aus und zieht empirische Untersuchungen und statistische Auswertungen heran (Koskinen 2020:9f.).

Einen anderen bzw. ergänzenden Ansatz bietet beispielsweise die Arbeit von Koskinen (2020). Sie umfasst „a view that emphasizes theorizing from social and cultural theories, [...] and aims at capturing the socio-cultural workings of affect” (Koskinen 2020:10). Sie möchte mit ihrem Buch einen Überblick über die möglichen Arten geben, auf die Affekte in der Translationswissenschaft⁸ untersucht werden können (Koskinen 2020:XI). Koskinen (2020) sieht Affekte dabei als „bridge concept, crossing over the various orientations in translation studies and also cutting through different contexts and modes of translatorial action” (Koskinen 2020:3). Sie hebt hervor, dass beispielsweise nicht nur die Übersetzungsstrukturen wie die unterschiedlichen Bereiche und institutionellen Hierarchien oder die Suche nach Übersetzungsgesetzen von Bedeutung sind, sondern auch die gelebte Realität dieser Tätigkeit und die involvierten Personen inklusive ihrer Bedürfnisse, Gewohnheiten u.Ä. wichtig sind (Koskinen 2020:X). Denn „to fully understand translation, one also needs to understand its *affective* side, the ways in which it forms a part of the lives of those involved with it” (Koskinen 2020:X).

Eine andere Herangehensweise stellt beispielsweise die Arbeit von Hubscher-Davidson (2018) dar, die sich mit Übersetzung und Emotion aus psychologischer Sicht beschäftigt. Dabei will sie jedoch die Emotionen der Übersetzer_innen in den Fokus rücken, da dieser Aspekt noch wenig Beachtung gefunden hat. Denn die Forschung hat sich zumeist mit der Übertragung von emotionaler Sprache auseinandergesetzt. Dabei spielen Emotionen in allen Arten von Problemlösungs- und Entscheidungsprozessen eine Rolle. Übersetzer_innen werden nicht nur vom emotionalen Gehalt des zu übersetzenden Materials, sondern auch von ihren eigenen Emotionen und jenen der Leser_innen in Ziel- und Ausgangssprache beeinflusst. Somit sind

⁸ Zwar trägt das Buch von Koskinen (2020) den Titel *Translation and Affect*, wobei das Dolmetschen nicht explizit beinhaltet ist, doch im Vorwort weist sie darauf hin, dass sie sowohl das Übersetzen als auch Dolmetschen abzudecken versucht (Koskinen 2020:XII).

drei Bereiche zentral: die Wahrnehmung, die Regulierung und der Ausdruck von Emotionen. So soll der Umgang der Übersetzer_innen mit diesen Emotionen betrachtet werden, indem auch emotionale Intelligenz thematisiert wird (Hubscher-Davidson 2018:2).

Allein an diesen zwei Beispielen ist zu erkennen, wie unterschiedlich an das Thema des Affekts in der Translationswissenschaft herangegangen werden kann. Während Hubscher-Davidson (2018) den Bereich Übersetzung und Emotion aus psychologischer Perspektive betrachtet, fokussiert sich Koskinen (2020) hingegen weniger auf neuropsychologische oder kognitive Aspekte, sondern mehr auf Aspekte sozialer und interaktioneller Art (Koskinen 2020:Xf.) Eine genauere Auseinandersetzung mit den unterschiedlichen Möglichkeiten, wie Affekte in der Translationswissenschaft untersucht werden können bzw. wie sie bereits untersucht und beschrieben wurden, würde den Rahmen dieser Arbeit sprengen. Daher seien zur Veranschaulichung lediglich einige willkürlich gewählte Beispiele und deren Forschungsinteresse im Bereich der Dolmetschwissenschaft genannt, um auf das breite Spektrum dieses Forschungsgebietes hinzuweisen.

Im Bereich des Dolmetschens hat sich zum Beispiel Furmanek (2006) mit Affekten beschäftigt und dabei hervorgehoben, welche Bedeutung Emotionen für die Tätigkeit des Dolmetschens haben. So wird durch die Konnotationen der Dolmetscher_innen, die emotionale Aspekte beinhalten, ein gewisser Hintergrund für ihre Arbeit gebildet, der beispielsweise die (womöglich unangemessene) Wahl der Termini für die Zielsprache beeinflusst, was also nicht unbedingt mit ihrer Berufserfahrung oder ihren sprachlichen Kompetenzen zusammenhängt (Furmanek 2006:67f.). Die Arbeit von den Ouden (2017) beispielsweise liefert einen anderen Blickwinkel. Sie beschäftigte sich mit dem Umgang von Dolmetscher_innen mit emotionalen Situationen. Dabei standen nicht die Emotionen der Dolmetscher_innen selbst, sondern die Emotionen Dritter im Fokus. Für den Bereich des *Community Interpreting* kann beispielsweise Cirillo (2010) genannt werden, die sich mit dem Thema der Affekte in gedolmetschten Gesprächen zwischen Ärzt_innen und Patient_innen beschäftigte. Einen anderen Fokus wählte Behr (2013a) mit ihrer Studie zum Einfluss der Stimmung auf die Beurteilung von Dolmetschleistungen. Aufgrund der Bedeutung ihrer Arbeit für das Forschungsinteresse der vorliegenden Masterarbeit soll auf die Studie separat eingegangen werden (siehe Kapitel 4.3.2.).

4.3. Zur Auswirkung von Stimmungen

Wie bereits erwähnt, wurden Affekte auch in der Psychologie lange außer Acht gelassen. Entsprechend stellt auch die Stimmungsforschung einen recht neuen Bereich dar und erst seit einigen Jahrzehnten rücken Stimmungen in den Fokus (Briner et al. 2000:22f.). Das Forschungsgebiet in der Psychologie, das sich mit den unterschiedlichen Stimmungsauswirkungen beschäftigt, ist sehr umfangreich. Mittlerweile herrscht Konsens darüber, dass Stimmungen auf unterschiedliche Art und Weise wirken können und einen Einfluss auf das Fühlen, Denken und Handeln haben können (Briner et al. 2000:93f.). Denn „feelings are part and parcel of how we think about life and their influence can be observed in any area of investigation” (Schwarz 2011:302). Somit unterliegen Bereiche wie das Verhalten, die Wahrnehmung, schlussfolgerndes Denken oder das Gedächtnis Einflüssen durch Stimmungen. Doch auch das Urteilen ist Stimmungseffekten unterworfen (Briner et al. 2000:93,115-133). In all diesen Fällen sind „jeweils unterschiedliche Kombinationen von Prozessen am Werk“ (Briner et al. 2000:93).

Prinzipiell kann zwischen drei Kategorien von Stimmungswirkungen unterschieden werden. Die erste Art stellen stimmungskongruente Effekte dar. Dabei beeinflusst die Stimmung psychologische Aspekte in einer Art und Weise, die der Stimmungstönung entspricht. Wenn Stimmungen einen gegenteiligen Effekt dessen, was man erwarten würde, bewirken, wird von Stimmungsinkongruenz gesprochen. Werden weder stimmungskongruente noch stimmungsinkongruente Aspekte untersucht, handelt es sich um nicht-evaluative Auswirkungen der Stimmungen. Ein sehr oft untersuchter Aspekt in diesem Kontext ist der Effekt auf die kognitive Verarbeitung. Hier fallen beispielsweise die Auswirkungen angenehmer und unangenehmer Stimmungen auf die Denkweise bzw. den Verarbeitungsstil hinein. Zusätzlich kann von direkten und indirekten, sowie bewussten und unbewussten Stimmungseinflüssen gesprochen werden. Während unbewusste Stimmungseffekte direkt sind und mit einer automatischen Verarbeitung einhergehen, sind es bei bewussten Stimmungen ihre kognitiven Repräsentationen, die einen Effekt bewirken, weshalb hier von indirekter Wirkung und einer kontrollierten Verarbeitung gesprochen werden kann (Briner et al. 2000:93-100).

Bis Mitte der 1970er Jahre untersuchte die experimentelle Stimmungsforschung insbesondere prosoziales Verhalten und beschäftigte sich kaum mit dem Stimmungseinfluss beim Urteilen (Clore & Schwarz 2003:296). Nach und nach rückte das Thema jedoch in den Fokus und bald wurden in diesem Bereich bedeutende Versuche von Clore & Schwarz (1983) durchgeführt (Behr 2013a:257). Die zwei Experimente stellen einen wichtigen Schritt bei der

Entwicklung der *Feelings-as-information theory* (siehe Kapitel 4.3.1.) dar und lieferten erste Erkenntnisse bzw. Nachweise für die Thesen der Theorie (Schwarz 2011:291f.). Die Theorie erläutert, welchen Einfluss Affekte, darunter Stimmungen, beim Urteilen haben können (Schwarz 2011).

Untersuchungen des Bewertungsvorgangs rückten die damit verbundenen internen Vorgänge bzw. kognitiven Prozesse der Rezipient_innen ins Forschungsinteresse. Auch in der dolmetschwissenschaftlichen Qualitätsforschung kam es zu einer Betonung dieser Aspekte, die beispielsweise aus einer soziologischen Perspektive betrachtet werden können (Behr 2013a:232). Wie bereits mehrfach angedeutet, setzt Behr (2013a) an dieser Stelle mit ihrer Studie an und untersucht ausgehend von der FAIT nach Schwarz (2010) den Einfluss von Stimmungen auf die Evaluation, wobei die rezipierende Person aus soziologischer Perspektive als die Person gesehen wird, die dem Gesagten den Sinn zuweist und deren Befindlichkeit somit eine Rolle bei der Qualitätsbeurteilung spielt (Behr 2013a).

Die FAIT nach Schwarz (2011) dient auch der vorliegenden Arbeit als Ausgangspunkt, weshalb nun genauer auf die Entstehung dieser Theorie eingegangen werden soll, bevor ihre Thesen näher beleuchtet werden. Da die Arbeit von Behr (2013a) die Inspiration für das Forschungsinteresse der vorliegenden Arbeit war, erscheint es sinnvoll, anschließend auch detaillierter auf ihre Studie einzugehen. Da die Studie von Behr (2013a) auf der FAIT basiert, soll diese direkt im Anschluss näher erläutert werden.

4.3.1. *Feelings-as-information theory* nach Schwarz (2011)

Bei der FAIT handelt es sich um eine Theorie, die die Rolle von Gefühlen beim Urteilen beschreibt. In dem Konzept werden unter dem Begriff Gefühle (*feelings*) dabei mehrere subjektive Erfahrungen zusammengefasst, die Teil des menschlichen Denkens bzw. Beurteilens sind und diese Prozesse beeinflussen. Darunter fallen unter anderem Emotionen und Stimmungen (Schwarz 2011:289).

4.3.1.1. Entstehung und Entwicklung

Die einfache persönliche Erfahrung bzw. Beobachtung, dass das Leben an sonnigen Tagen besser aussieht als an verregneten weckte bei Norbert Schwarz das Interesse an dieser Thematik und veranlasste ihn dazu, sich mit der experimentellen Stimmungsforschung auseinanderzusetzen. Dies brachte ihn von der Universität in Mannheim an die University of Illinois in die USA und führte schließlich zu den ersten zwei Experimenten, auf die die

Feelings-as-information theory zurückgeht. Die Ergebnisse lieferten erste Nachweise für einige Annahmen, die später zu Kernthemen bei der Entwicklung der FAIT wurden (Schwarz 2011:290ff.). Diese ersten Versuche wurden in den Jahren 1980-1981 von Gerald L. Clore und Norbert Schwarz durchgeführt, als letzterer ein Postdoc an der University of Illinois at Urbana-Champaign war (Clore & Schwarz 2003:296).

Anfangs lag der Fokus auf Stimmungen und auf dem Einfluss von positiven bzw. negativen Stimmungen auf evaluative Urteile. Aus diesem Grund war zunächst von *mood-as-information* die Rede. Aber es zeigte sich, dass die theoretischen Prinzipien ebenfalls auf andere Arten von Gefühlen übertragen und zu einem umfassenderen Konzept erweitert werden konnten, das das Zusammenspiel von Denken und Fühlen beschreibt (Schwarz 2011:289). So zeigten spätere Studien, dass die informative Funktion auch auf metakognitive Erfahrungen, wie die Leichtigkeit beim Abrufen, und auch auf körperliche Empfindungen erweitert werden kann (Schwarz 2011:292). Schwarz (2011) spricht in seiner Theorie von „subjective experiences, including moods and emotions, metacognitive feelings [...] and bodily sensations” (Schwarz 2011:289).

Die FAIT hat somit mehrere Phasen durchlaufen und besteht nun in ihrer dritten Fassung. Im Gegensatz zu den vorangegangenen Versionen betont die aktuelle Theorie die Wichtigkeit der Fragestellung, auf die ein Gefühl bezogen wird. Auch die Bandbreite der möglichen Einflussfaktoren auf den Verarbeitungsstil wurde erweitert, während die Rolle der Verarbeitungsfähigkeit nun als geringer angesehen wird (Schwarz 2011:293). Auf die genauen Grundzüge der FAIT soll nun im Folgenden eingegangen werden.

4.3.1.2. Grundzüge der Theorie

Die *Feelings-as-information theory* nach Schwarz (2011) basiert auf vier Postulaten bzw. Hauptthesen, die jeweils weiter ausgeführt werden. Diese Annahmen beziehen sich auf mehrere Aspekte von Gefühlen. Zum einen beschreiben sie, wie der Informationswert von Gefühlen wahrgenommen wird und auf welche Art und Weise diese Gefühle als Grundlage beim Fällen von Urteilen genutzt werden. Schließlich beschreiben die Postulate auch, wie sich Gefühle auf den Verarbeitungsstil auswirken, der in einer Situation spontan angenommen wird (Schwarz 2011:293)

Die erste These der FAIT besagt, dass Gefühle als Informationsquelle dienen und beschäftigt sich damit, was Gefühle einer Person vermitteln. Menschen greifen beim Bilden von Urteilen auf ihre Gefühle zurück und stellen sich die Frage, wie sie sich dabei fühlen. Die

unterschiedlichen Arten von Gefühlen liefern dabei auch unterschiedliche Arten von Informationen (Schwarz 2011:292ff.). So haben Emotionen, wie auch oben bereits erwähnt, eine erkennbare Ursache bzw. einen identifizierbaren Bezugspunkt, auf den sie sich beziehen. Bei Stimmungen hingegen fehlt dieser Bezug. Sie setzen graduell ein und können für einen längeren Zeitraum bestehen bleiben (Schwarz 2011:294). Daher sind sie „more diffuse than emotions and primarily convey generic valence information that lacks a clear referent” (Schwarz 2011:294). Schließlich beschreibt Schwarz (2011) auch *cognitive feelings* wie beispielsweise Langeweile oder Überraschung, die Informationen in Hinblick auf den Wissensstand liefern, und *bodily experiences* wie Schmerz oder Hunger, die uns über den physischen Zustand unseres Körpers informieren (Schwarz 2011:294f.).

Der zweiten These zufolge hängt der Einfluss von Gefühlen davon ab, wie ihr Informationswert in Hinblick auf die durchzuführende Aufgabe wahrgenommen wird, also welchen informativen Wert eine Person ihren Gefühlen beimisst. Dabei kann zwischen Gefühlen unterschieden werden, die vom zu beurteilenden Objekt hervorgerufen werden (*integral feelings*) und solchen, die ihren Ursprung woanders haben (*incidental feelings*) und somit potenziell irreführende Informationen liefern können. Menschen nehmen zwar ihre Gefühle wahr, sind jedoch wenig sensibel dafür, woher sie kommen. So werden im Normalfall ein Gefühlszustand und damit verbundene Gedanken auf das Objekt bezogen, das in dem Moment im Fokus steht. Schließlich gäbe es sonst keinen Grund, dass man diese Gefühle in diesem bestimmten Kontext empfinden sollte (Schwarz 2011:295). Dabei gilt, dass

the use of feelings as a source of information does not require a conscious attribution of the feeling to the target, [...] [j]ust as we use declarative information that happens to come to mind, we also use experiential information that happens to come to mind. In either case, we assume that the information is relevant to what we think about, or else why would it come to mind? (Clore & Schwarz 2003:299)

Menschen tendieren also dazu „to perceive incidental feelings as being ‘about’ the target of judgment” (Schwarz 2011:295). Dies bleibt solange bestehen, bis der Gefühlszustand auf eine andere Quelle bezogen wird, wobei es aber keine Rolle spielt, ob diese Zuschreibung korrekterweise erfolgt oder nicht (Schwarz 2011:295). Denn „[w]hen people (correctly or incorrectly) attribute their feelings to an incidental source, the perceived informational value of their feelings for the judgment at hand is undermined” (Schwarz 2011:295). Wenn Menschen nun beginnen, den informativen Wert ihrer empfundenen Gefühle zu hinterfragen, suchen sie nach anderen Informationsquellen, die sie als Grundlage ihres Urteils heranziehen können. Sind

solche vorhanden, wird auf sie zurückgegriffen und folglich sinkt der Gefühlseinfluss in so einer Situation. Stellen Menschen jedoch fest, dass sie diese Gefühle trotz möglicher anderer Erklärungen haben, verstärkt dies den wahrgenommenen Informationswert. Können keine anderen Informationen herangezogen werden, kommt es beim Urteilen zu einer Verzerrung in die entgegengesetzte Richtung, die durch eine Überkorrektur zustande kommt. Ferner geht die Theorie im Rahmen der zweiten These auch davon aus, dass Veränderungen der Gefühle informativer sind als stabile Gefühlszustände (Schwarz 2011:295).

Drittens postuliert Schwarz (2011:296f.), dass wenn Menschen ihre Gefühle als Informationsquelle heranziehen, dies nach den denselben Prinzipien geschieht, wie bei anderen Informationen. Doch nur wenn der informative Wert nicht hinterfragt wird, werden Gefühle als Quelle herangezogen. Ihr Einfluss steigt, wenn sie als relevant für die zu fällende Beurteilung eingestuft werden. Sobald jedoch andere relevante Informationen vorhanden sind, sinkt der Gefühlseinfluss. Somit spielt in diesem Kontext auch die Expertise eine Rolle. Denn „people are less likely to rely on their feelings when they have high expertise in the domain of judgment [...] which presumably facilitates the assessment of the relevance of one’s feelings and renders other information easily accessible” (Schwarz 2011:297). Weiters wird diese Suche nach alternativen Informationen und die Einschätzung der Relevanz von empfundenen Gefühlen in Situationen mit geringer Motivation und geringer Verarbeitungskapazität erschwert, weshalb die Auswirkungen von Gefühlen dann stärker sind. Schließlich spielt auch die spezifische Fragestellung eine entscheidende Rolle (Schwarz 2011:297). Während die meisten Forschungen mehr positive Urteile bei positiver Stimmung nachweisen konnten, zeigten andere Studien auch auf, dass positive Gefühle auch negative Urteile zur Folge haben können. Dies geschieht beispielsweise, wenn wir bei der Lektüre einer traurigen Geschichte positive Gefühle empfinden und daraufhin schlussfolgern, dass es sich um eine schlechte traurige Geschichte handelt, da wir ansonsten ein trauriges Gefühl wahrgenommen hätten. Somit hängt der Gefühlseinfluss einer Person auch von der spezifischen Frage ab, auf die das Gefühl angewendet wird (Schwarz 2011:292). Die unterschiedlichen Fragestellungen können weiters auch zu unterschiedlichen Verhaltensweisen bei gleicher Stimmung führen, je nachdem auf welche Frage sie angewendet wird. Es konnte beispielsweise beobachtet werden, wie positiv gestimmte Personen bei der Frage, ob sie mit ihrer Leistung bei einer Aufgabe zufrieden waren, die Aufgabe beendeten, während negativ gestimmte Personen die Aufgabe fortführten. Umgekehrt führten positiv gestimmte Personen bei der Frage, ob ihnen die Aufgabe Spaß mache, fort, während negativ gestimmte Personen bei dieser zweiten Frage die Tätigkeit beendeten. In beiden Fällen verhielten sich die Personen gemäß Informationen, die ihnen ihre

Stimmung lieferte. Dies führte jedoch aufgrund der unterschiedlichen Fragestellungen zu anderen Verhaltensimplikationen. Somit sind Schlussfolgerungen aus Gefühlen formbar und kontextsensitiv. Wichtig ist an dieser Stelle noch einmal darauf hinzuweisen, dass verschiedene Gefühle auch ein unterschiedliches Maß an Interpretation erfordern und einige von ihnen eine größere Bandbreite an möglichen Schlussfolgerungen zulassen, wie dies beispielsweise bei Stimmungen der Fall ist (Schwarz 2011:297).

Die vierte und letzte These der FAIT besagt, dass Gefühle nicht nur als Urteilsgrundlage dienen, sondern auch die Art und Weise der Informationsverarbeitung beeinflussen. Dabei wird davon ausgegangen, dass kognitive Prozesse auf die Umwelt abgestimmt werden und so beispielsweise die passende Verarbeitungsstrategie gewählt wird. In Situationen, in denen alles reibungslos abläuft, verlassen sich Menschen auf Routinen und vorhandene Wissensstrukturen, die in der Vergangenheit erfolgreich waren. Dies geht gleichzeitig mit einer Bereitschaft einher, auch Risiken einzugehen. In problematischen Situationen hingegen verlassen Menschen sich nicht länger auf altbewährte Routinen, sondern konzentrieren sich verstärkt auf die konkrete Situation. Auch Gefühle können uns an dieser Stelle schnell und einfach signalisieren, in was für einer Situation wir uns befinden und beeinflussen somit diesen Prozess (Schwarz 2011:298). Positive Stimmungen begünstigen dabei einen Top-down-Verarbeitungsstil, der auf allgemeinen Wissensstrukturen beruht. Dies geht mit weniger Fokus und einer höheren Kreativität einher. Negative Stimmungen auf der anderen Seite fördern einen systematischen Bottom-up-Verarbeitungsstil, der sich durch einen Fokus auf Details und eine begrenzte Kreativität auszeichnet (Schwarz 2011:293). Dies bedeutet jedoch nicht, dass positiv gestimmte Menschen nicht fähig sind, analytisch zu verarbeiten oder dies nicht wollen. Ihre Gefühle vermitteln lediglich keine Notwendigkeit, dies zu tun. Wenn die Anforderungen einer Situation es zulassen, keine expliziten Anweisungen es verbieten und beide Verarbeitungsarten zur Erreichung der Ziele angewendet werden können, beeinflussen also die Stimmungen die spontane Annahme des jeweiligen Verarbeitungsstils. Welcher Verarbeitungsstil in der jeweiligen Situation dabei vorteilhafter wäre, hängt von der spezifischen Aufgabe ab (Schwarz 2011:302ff.).

4.3.2. Einfluss von Stimmungen auf die Evaluation - Studie von Martina Behr (2013a)

Qualitätsbewertung im Dolmetschen kann, wie bereits dargelegt, zum Beispiel aus der Sicht der Rezipient_innen erfolgen (siehe Kapitel 2.3.2. bzw. 2.3.2.2.). In ihrer Arbeit verbindet Behr (2013a) Qualitätsevaluation mit dem Aspekt der Stimmung und stellt so das Befinden der

Rezipient_innen in den Vordergrund. Sie wählt einen interdisziplinären Blick und bezieht sich auf Erkenntnisse aus den Bereichen der Soziologie, Sozialpsychologie und der Kommunikationswissenschaft. Eine zentrale Theorie, auf die sie sich bezieht, ist die FAIT nach Schwarz (2010). Behr (2013a) ging der Frage nach, inwiefern die Beurteilung einer Dolmetschung durch dieselbe Person – zu zwei unterschiedlichen Zeitpunkten durchgeführt – mit ihrer Stimmung korreliert werden kann. Dabei ist den Teilnehmer_innen nicht bewusst, dass sie zweimal dieselbe Leistung evaluieren. Um dies zu beantworten, wurde ein Versuch mit zwei Gruppen durchgeführt. Die erste umfasste elf erfahrene Konferenzdolmetscher_innen mit Deutsch als A-Sprache und Französisch als B- oder C-Sprache. Sie sollten die Leistungen so bewerten, als ob sie eine Dolmetschung im Rahmen einer Abschlussprüfung evaluieren würden. Eine zweite Gruppe wurde aus zwölf Medizinstudent_innen gebildet und nahm die Perspektive von User_innen ein (Behr 2013a:273f.).

Beide Versuchsgruppen sollten an zwei auseinander liegenden Terminen, die im Abstand von etwa 8 Wochen geplant wurden, aufgezeichnete Verdolmetschungen, in denen französische Ausgangstexte ins Deutsche gedolmetscht wurden, anhören und bewerten. Dabei wurde den Teilnehmer_innen, ohne dass diese es wussten, zu den unterschiedlichen Terminen zwei Mal eine gleiche Dolmetschung als dritte Aufnahme zur Bewertung vorgespielt. Am Ende der jeweiligen Termine füllten die Zuhörer_innen auch einen Stimmungsfragebogen aus. Die Termine wurden für die beiden Gruppen jeweils getrennt voneinander durchgeführt, wobei auch die Versuchsanordnung Unterschiede aufwies. Neben der Anzahl der eingespielten Aufnahmen erhielten die Gruppen auch teils unterschiedliche Unterlagen. Die Gruppe der Prüfer_innen hatte beispielsweise auch eine Kopie des Originaltextes beim Anhören zur Verfügung (Behr 2013a:273f.,277ff.). Während beide Gruppen denselben Stimmungsfragebogen erhielten, wurden ihnen für die Evaluation der Verdolmetschungen verschiedene Fragebögen vorgelegt. Jener der Konferenzdolmetscher_innen umfasste eine Auflistung von den Qualitätsparametern Vollständigkeit, Richtigkeit, Funktionalität, Grammatik/Syntax, Ausdruck und Prosodie, die es auf einer Skala von hervorragend bis ungenügend zu bewerten galt. Zudem wurde eine Gesamtnote vergeben. Die Gruppe der Student_innen erhielt eine Information in Hinblick darauf, was eine qualitativ gute Dolmetschung auszeichnet. Der Bewertungsbogen war im Gegensatz zum ersten allgemeiner ausformuliert. Gefragt wurde beispielsweise danach, wie leicht oder schwer das Zuhören gefallen ist, oder als wie angenehm die Dolmetschung empfunden wurde. Doch auch hier wurde nach einer Gesamtnote gefragt und die Leistung sollte zusätzlich auch auf einer Skala von hervorragend bis sehr schlecht eingestuft werden. Bei drei der Fragen war weiters Platz für eine Begründung (Behr 2013a:343ff.,348).

Behr (2013a) stellte für ihren Versuch zwei Hypothesen auf, die auf den Postulaten der FAIT basierten. Sie ging davon aus, dass alle bzw. die Mehrheit der Teilnehmer_innen die zweimal präsentierte Dolmetschung anders bewerten würde. Zweitens nahm sie an, dass die Unterschiede in den Bewertungen sich in den Stimmungsangaben widerspiegeln würden (Behr 2013a:279f.). Die Auswertung ergab, dass in beiden untersuchten Gruppen die Mehrheit der Teilnehmer_innen die zwei Aufnahmen zu den zwei Terminen unterschiedlich bewertete. Unter den Studierenden vergaben drei Personen die gleiche Note, während dies von den Dolmetscher_innen nur eine Person tat. So konnte ihre erste Hypothese für beide Gruppen bestätigt werden (Behr 2013a:292ff.). Im nächsten Schritt wurde ein Blick auf eine potenzielle Korrelation zwischen Noten und Stimmungen geworfen. Denn wenn Stimmungen bei komplexeren Aufgaben gemäß der FAIT als Informationsquelle dienen, könnten Unterschiede in der Stimmung mit der Vergabe der unterschiedlichen Noten zusammenhängen. Tatsächlich konnte sie in beiden Gruppen Unterschiede in der Stimmung feststellen. Dennoch waren fast alle Teilnehmer_innen an beiden Versuchsterminen positiver Stimmung. Bei der Gegenüberstellung der Notendifferenz und der Stimmungsdifferenz war in keiner der Gruppen eine klare Korrelation zu erkennen (Behr 2013a:294-299). So war „keine einheitliche Tendenz in dem Sinne festzustellen, dass z.B. alle weniger/stärker gut gestimmten Probanden tendenziell eher positive/negative Noten vergeben“ (Behr 2013a:297). Auch die Betrachtung der Prüfer_innen, die sowohl in negative als auch positive Richtung jeweils die größte Differenz bei den erhobenen Stimmungen aufwiesen, ergab kein bestimmtes Bild (Behr 2013a:297f.) und so „variieren die Stimmungsänderungen nicht in Einklang mit den Notendifferenzen“ (Behr 2013a:297). So konnten zwar sowohl in der Stimmung als auch in der Vergabe der Noten Variationen aufgezeigt werden, eine Korrelation, bei der die Stimmung die Beurteilung tendenziell in eine bestimmte Richtung beeinflusst, konnte im Rahmen von Behrs Versuch nicht aufgezeigt werden. Sie weist darauf hin, dass dies möglicherweise der Tatsache geschuldet ist, dass die Mehrheit der Teilnehmer_innen an beiden Versuchstagen positiver Stimmung waren (Behr 2013a:297f.). Im nächsten Schritt sollte überprüft werden, ob eine getrennte Betrachtung der zwei Stimmungsdimensionen eine Tendenz erkennen ließe. So stellte Behr die Differenzen der vergebenen Gesamtnoten für beide Gruppen jeweils den Dimensionen der Aktivität und Gestimmtheit gegenüber. Jedoch ergab auch dies bei keiner der beiden Gruppen für keine der zwei Dimensionen ein klares Bild (Behr 2013a:299ff.).

Ein Blick auf die Beurteilung der einzelnen Aspekte ergab, dass die Evaluierung der zwei kommunikativ-sprachlichen Qualitätskriterien „Ausdruck“ und „Funktionalität“ stärkere Schwankungen aufwiesen als die inhaltsbezogenen Parameter, die eher objektiver fassbar sind

und auf einer weniger subjektiven Beurteilungsgrundlage basieren. So ist es transparenter und leichter zu erläutern, ob eine Äquivalenz zwischen Ausgangs- und Zieltext gegeben ist, oder ob Fehler in der Grammatik oder Syntax gemacht wurden, als die kommunikativ-sprachlichen Aspekte zu bewerten, die gemäß der FAIT eher von Schwankungen in der Stimmung beeinflusst werden. Wenn der Fokus hingegen auf inhaltlichen Parametern liegt, scheint die Stimmung weniger Einfluss zu nehmen. Behr wies jedoch darauf hin, dass die geringe Schwankung bei der Prosodie überraschend war (Behr 2013a:303f). Denn

[g]erade Tonhöhe, Klang und Modulation der Stimme sind sicherlich Bereiche, denen unweigerlich mit individuellem Empfinden entgegengetreten wird – das heißt gerade hier wäre aufgrund des Fehlens der objektiven Bewertungskriterien das Hinzuziehen der Stimmung als Beurteilungsgrundlage zu erwarten. (Behr 2013a:304)

Eine mögliche Erklärung für dieses Ergebnis wäre, dass den Konferenzdolmetscher_innen der subjektive Aspekt beim Bewerten dieses Parameters bewusst ist, und sie ihm daher weniger Wichtigkeit beimessen (Behr 2013a:304).

Behr (2013a) setzte im Zuge ihrer Auswertung auch die Notenvergabe und die Erfahrung ihrer Teilnehmer_innen als Konferenzdolmetscher_innen und Prüfer_innen in Relation zueinander, konnte jedoch keine Tendenz aufzeigen. So merkte sie an, dass Dolmetscher_innen womöglich schon in ihrer Ausbildung genügend Expertise erlangen, um sich beim Evaluieren von Dolmetschleistungen nicht von ihren Stimmungen beeinflussen zu lassen. Denn um professionell zu dolmetschen, müssen Studierende lernen, ihre Empfindungen auszublenden. Dies könnte auch der Grund sein, warum sich die Dolmetscher_innen in der Studie nicht von Affekten beeinflussen haben lassen. An dieser Stelle merkte Behr an, dass eine Untersuchung von Studierenden am Anfang des Studiums und von Examenskandidat_innen mehr Aufschluss geben könnte (Behr 2013a:301f.). Basierend auf diesen Feststellungen von Behr (2013a) wurde der Grad der Expertise explizit in die gestellten Forschungsfragen der vorliegenden Arbeit aufgenommen.

5. Methode, Versuchsdesign und Durchführung

Die Arbeit zum Stimmungseinfluss auf die Evaluation im Bereich des Simultandolmetschens von Martina Behr (2013a) (siehe Kapitel 4.3.2.) war Inspiration für die vorliegende Arbeit. Aus den Schlussfolgerungen der Studie entstand das Vorhaben, einen empirischen Versuch am Zentrum für Translationswissenschaft in Wien durchzuführen. Dieser sollte den Einfluss der Stimmung auf die Bewertung von Simultandolmetschleistungen durch Studierende im Bereich des Konferenzdolmetschens untersuchen, wobei auch der Grad an Expertise mit einbezogen werden sollte. Basierend auf den Erläuterungen in den vorangegangenen Kapiteln soll an dieser Stelle nochmals festgehalten werden, dass die vorliegende Arbeit von Qualität als soziales Konstrukt ausgeht (siehe Kapitel 2.2.) und basierend auf Behr (2013a) die zentrale Rolle der Rezipient_innen (siehe hierzu Kapitel 2.3.2.2.) in den Vordergrund stellt. Die Arbeit ist des Weiteren im didaktischen Bereich zu verorten und beinhaltet eine Evaluation durch Studierende in einer Form, die als *peer feedback* (siehe Kapitel 2.3.3.) bezeichnet werden kann und im Zuge des empirischen Versuchs durchgeführt wurde. Die Erhebung der Stimmung und die hierzu behandelten Erkenntnisse orientieren sich an den Erläuterungen und der Herangehensweise von Behr (2013a). Dementsprechend basiert die vorliegende Arbeit, wie bereits erwähnt, auch auf der FAIT (siehe Kapitel 4.3.1.).

5.1. Methode und Versuchsdesign

Die Dolmetschwissenschaft zeichnet sich durch eine Vielfalt an methodischen Traditionen und Perspektiven aus, die jeweils einen wichtigen Beitrag für das Forschungsfeld leisten (Liu 2011:107). Dabei stützen sich die unterschiedlichen Forschungsansätze vor allem auf die sozialwissenschaftlichen Disziplinen Anthropologie, Soziologie und Linguistik. Aber auch Methoden aus beispielsweise der Psychologie kommen zum Einsatz und es wird experimentelle Forschung betrieben. In den letzten Jahrzehnten ist ein Anstieg an experimentellen Studien im Dolmetschen zu verzeichnen. Da solche Studien oftmals in der Sprachforschung durchgeführt werden, eignen sie sich auch für den Bereich des Dolmetschens und kommen typischerweise im Simultan- bzw. Konferenzdolmetschen zum Einsatz (Hale & Napier 2013:149f.).

Seit den Anfängen der Qualitätsforschung durch die Psychologie wurde in der empirischen, experimentellen Forschung untersucht, welche unabhängigen Variablen die abhängige Variable der Qualität, die implizit in der Dolmetschperformenz enthalten war,

beeinflussten (Pöchhacker 2013:37). So wurde oftmals der Einfluss unabhängiger Variablen wie der Sprechgeschwindigkeit, Dichte oder Müdigkeit untersucht (Gile 2003:110). Liu (2011) untersuchte 48 Studien zum Thema Dolmetschen, die in der Fachzeitschrift *Interpreting*⁹ in den zwölf Ausgaben im Zeitraum von 2004 bis 2005 veröffentlicht wurden. Die analysierten evidenzbasierten Studien wurden in Hinblick auf ihren Ansatz bzw. die Methode, die analysierten Variablen und die Datenerhebungs- sowie Auswertungsverfahren analysiert (Liu 2011). Sie stellte bei ihrer Sichtung fest, dass Versuche oftmals als Experiment bezeichnet werden, ohne tatsächlich dessen Merkmale zu erfüllen (Liu 2011:107).

Um als Experiment zu gelten, muss ein Versuch die Manipulation von einer oder mehreren Variablen beinhalten, sowie die Messung der Effekte dieser Manipulation auf das Verhalten umfassen (Liu 2011:100). Ziel hierbei „ist die theoriegestützte Überprüfung des kausalen Einflusses einer oder mehrerer **unabhängiger Variablen** [...] auf die Ausprägungen einer oder mehrerer **abhängiger Variablen**“ (Döring 2023:196). Dabei stellen unabhängige Variablen die untersuchten Ursachefaktoren dar, während mittels abhängiger Variablen der Effekt bzw. die Wirkung gemessen wird (Döring 2023:196). Zusätzlich zeichnet sich ein Experiment durch die Bildung zweier Gruppen, der Experimental- und Kontrollgruppe, aus. Die Zuordnung der Proband_innen zu den beiden Gruppen erfolgt dabei zufällig, sprich randomisiert (Döring 2023:195; Hale & Napier 2013:158). Jede teilnehmende Person hat dabei die gleiche Wahrscheinlichkeit, in eine der beiden Gruppen eingeteilt zu werden (Hale & Napier 2013:158). Eine randomisierte Zuteilung ist jedoch nicht immer möglich. Dies kann beispielsweise mit der Größe der Grundgesamtheit oder der zu untersuchenden Variablen zusammenhängen (Hale & Napier 2013:158). Schließlich werden die unabhängigen Variablen, auch experimentelle Variablen genannt, im Rahmen eines Experiments manipuliert (Döring 2023:196).

Liu (2011) stellte also fest, dass manche der von ihr betrachteten dolmetschwissenschaftlichen Studien als Quasi-Experimente oder Prä-Experimente bezeichnet werden müssen (Liu 2011:102f.). Auch wenn experimentelle Designs am besten geeignet sind, um Ursache und Wirkung zu untersuchen (Döring 2023:196; Liu 2011:108), gilt, dass „quasi-experiments, pre-experiments, and even descriptive studies, if done properly, can greatly enhance our understanding of the phenomena in our field“ (Liu 2011:108). Gemäß der von Liu

⁹ Hierbei handelt es sich um eine erstmals 1996 herausgegebene Fachzeitschrift, die sich der dolmetschwissenschaftlichen Forschung widmet und hauptsächlich evidenzbasierte Studien veröffentlicht. Deshalb war sie nach der Ansicht Lius auch für das Review zu den aktuellen Entwicklungen in Hinblick auf verwendete Methoden gut geeignet (Liu 2011:86).

(2011) aufgestellten Auflistung der vorgefundenen Studiendesigns fällt der empirische Versuch im Rahmen dieser Arbeit am ehesten in die Kategorie des Quasi-Experiments. Diese Form kommt zum Einsatz, wenn keine randomisierte Zuteilung in Kontroll- und Experimentalgruppe möglich ist (Liu 2011:102). Stattdessen wird mit bereits vorgefundenen Gruppen gearbeitet (Döring ⁶2023:201).

Wie bereits erläutert, setzt die vorliegende Arbeit an der Studie von Behr (2013a) an. Auch der empirische Versuch, der im Rahmen dieses Vorhabens entwickelt wurde, basiert auf jenem von Behr (2013a) und orientiert sich an deren Aufbau und Verlauf. Ziel der hier vorliegenden Arbeit war es zu untersuchen, wie sich die Bewertungen einer ausgewählten Dolmetschaufnahme im Rahmen eines Peerfeedbacks durch Studierende an zwei verschiedenen Terminen unterscheiden und welche Auswirkung die Stimmung der Studierenden hierbei hat. Dabei wurden die Teilnehmenden nicht darüber informiert, dass sie eine bestimmte Aufnahme an beiden Terminen zu hören bekamen. Zusätzlich war, basierend auf den Annahmen der FAIT (siehe Kapitel 4.3.1.), auch die Rolle der Expertise in Hinblick auf den potenziellen Stimmungseinfluss von Interesse. So wurden die Studierenden bereits im Vorhinein basierend auf dem Grad der Expertise gezielt ausgewählt und in zwei Gruppen eingeteilt. Somit kann die Stimmung der Personen als eine unabhängige Variable definiert werden, die sich auf die Bewertung auswirkt. Eine zusätzliche Variable stellt auch der Grad der Expertise dar. Diese hat gemäß der FAIT nach Schwarz (2011) womöglich Einfluss darauf, ob die teilnehmenden Personen sich von ihrer Stimmung leiten lassen. Eine randomisierte Zuteilung in zwei Gruppen ist somit aufgrund des Forschungsinteresses der vorliegenden Arbeit nicht möglich. Wie bereits angedeutet, ist dies nicht immer möglich

due to constraints of the context that already exist. This is referred to as quasi-experimental research because not all of the variables can be completely controlled. Aspects of each group may be the same and testing takes place under the same conditions, but there are some variables that are different. (Hale & Napier 2013:159)

An dieser Stelle muss jedoch festgehalten werden, dass das Forschungsdesign nicht gänzlich einem Quasi-Experiment entspricht. Denn „[a]uch im Quasi-Experiment wird die unabhängige Variable aktiv variiert, indem den Untersuchungsgruppen jeweils eine Behandlung (Treatment) zukommt, während die Kontrollgruppe unbehandelt bleibt“ (Döring ⁶2023:201). Im empirischen Versuch dieser Arbeit kann nicht von einer Versuchs- und Kontrollgruppe im klassischen Sinne gesprochen werden. Der zeitliche Abstand zwischen den Terminen ist gegeben, was womöglich eine Änderung in der Stimmung mit sich bringt. Trotzdem wurde im

geplanten Versuch keine gezielte Manipulation vorgenommen. Die Stimmung wird zwar gemessen und ihr Einfluss untersucht, sie wird jedoch nicht gewollt verändert bzw. beeinflusst.

Im Folgenden sollen das Forschungsinteresse und die Forschungsfragen der vorliegenden Arbeit im Detail erläutert werden, bevor auf die die zwei Teilnehmer_innen-Gruppen eingegangen wird, gefolgt von einer detaillierten Beschreibung des Versuchsablaufs.

5.1.1. Ausgangspunkt und Forschungsfragen

Basierend auf der Arbeit von Behr (2013a) wurde die erste Forschungsfrage der vorliegenden Arbeit folgendermaßen formuliert:

- Wie unterscheiden sich die Bewertungen derselben Simultandolmetschung, die unwissentlich zu zwei unterschiedlichen Zeitpunkten durch Studierende vorgenommen werden, und welche Rolle kommt hierbei der Stimmung zu?

Ausgehend vom Konzept der Qualität als soziales Konstrukt (siehe Kapitel 2.2.) und der angenommenen Bedeutung der Zuhörer_innen (siehe Kapitel 2.3.2.2.) sowie der damit verbundenen Auffassungen und der Arbeit von Behr (2013a), wurde davon ausgegangen¹⁰, dass die Bewertungen der untersuchten Aufnahme unterschiedlich ausfallen würden. Dabei wurde einerseits erwartet, dass es zu Unterschieden zwischen den Teilnehmer_innen in Hinblick auf die vergebenen Noten kommen würde. Andererseits wurde angenommen, dass auch jeweils dieselbe Person die gleiche Aufnahme zu unterschiedlichen Zeitpunkten potenziell anders bewerten könnte.

Wie bereits erläutert wurde, geht die FAIT nach Schwarz (2011) davon aus, dass Personen weniger auf Stimmungen als Informationsquelle zurückgreifen, wenn sie über Expertise in dem Bereich, in dem sie urteilen sollen, verfügen (siehe Kapitel 4.3.1.2.). Bei ihrer Auswertung stellte Behr (2013a) fest, dass die Erfahrung als Konferenzdolmetscher_in bzw. Prüfer_in in keinem Zusammenhang mit der Vergabe der Noten stand. So verwies sie auf die Expertise und die Möglichkeit, dass diese bereits in der Ausbildung in ausreichendem Maße erlangt wird (Behr 2013a:301f.). Ausgehend von diesen Schlussfolgerungen und basierend auf den Ausführungen in Kapitel 3 wird Expertise in der vorliegenden empirischen Untersuchung

¹⁰ An dieser Stelle wurden gezielt keine klassischen Hypothesen formuliert, da es sich beim vorliegenden Versuch auch um kein klassisches Experiment im eigentlichen Sinne handelt. Stattdessen wird basierend auf der aufbereiteten Theorie und der Arbeit von Behr (2013a) von Annahmen gesprochen.

durch Erfahrung und Übung in der Ausbildung definiert, die zur Erlangung bestimmter Kompetenzen im Dolmetschbereich führen und eine Unterscheidung zwischen Noviz_innen und Expert_innen ermöglichen. Als Meilenstein, der einen Übergang zum Expert_innen-Status darstellt, wurde die positiv bestandene Modulprüfung am ZTW definiert. So wurden zwei Gruppen für den Versuch eingeplant, die nach dem Fortschritt im Studium zusammengestellt wurden. Dementsprechend lautete die zweite Frage wie folgt:

- Welche Unterschiede ergeben sich aufgrund des Studienfortschritts bzw. aufgrund des unterschiedlichen Grades an Expertise der bewertenden Personen?

Wie bereits beschrieben, ergab die Auswertung bei Behr (2013a) zudem, dass es im Vergleich zu inhaltsbezogenen Parametern bei der Evaluierung zu mehr Schwankungen bei den zwei Qualitätskriterien „Ausdruck“ und „Funktionalität“ kam. Diese Erkenntnisse bzw. Schlussfolgerungen waren ein Anstoß, im Rahmen des empirischen Versuchs auch den potenziellen Unterschied bei der Bewertung der unterschiedlichen Einzelaspekte zu ermitteln. So wurde auch der nachstehenden Forschungsfrage nachgegangen:

- Inwieweit unterscheiden sich die Bewertungen der beiden Gruppen in Hinblick auf die inhaltsbezogenen bzw. formalen Qualitätsparameter?

Um den gestellten Fragen nachzugehen, wurde zunächst nach passenden Personen für den empirischen Versuch gesucht.

5.1.2. Beschreibung der Versuchsteilnehmer_innen

Für den Versuch wurden zwei Gruppen aus Studierenden am ZTW zusammengestellt. Hierfür wurden gezielt Personen gesucht, die sich in zwei Gruppen je nach Studienfortschritt bzw. Expertise einteilen ließen. Hinzu kamen die Voraussetzungskriterien Englisch als A- oder B-Sprache und Schwerpunkt Dolmetschen. Über einen längeren Zeitraum von etwa drei Monaten wurde versucht, Personen auf verschiedene Art und Weise zu erreichen und für den Versuch zu gewinnen. Dabei wurden Informationen und Bitten zur Teilnahme über Kolleg_innen weitergeleitet. Außerdem wurden Aufrufe in den Online-Gruppen und Chats, die seitens der Studierenden am ZTW gebildet und organisiert wurden, verfasst. Zudem wurden einige Lehrpersonen, die u.a. Lehrveranstaltungen im ersten Semester leiten, kontaktiert, die einen

Aufruf und Informationen zum geplanten Versuch an Studierende ihrer Lehrveranstaltungen weiterleiteten. Trotz dieser zahlreichen Bemühungen konnten für den Versuch nicht mehr als insgesamt 8 Teilnehmer_innen gewonnen werden, die sinnvoll in zwei Gruppen eingeteilt werden konnten¹¹. Dies scheint jedoch nicht allzu überraschend, wenn man bedenkt, dass „[a] small sample size has always been a problem in interpreting research. Very often the size of the participant pool is not large enough to achieve significant results in statistics” (Liu 2011:107). Dennoch können gewiss Tendenzen erhoben und beschrieben werden, sowie mögliche noch zu erforschende Aspekte in dem Bereich aufgezeigt werden.

Schließlich umfasste der Versuch zwei Gruppen aus jeweils vier Personen. Die erste Gruppe bestand aus Studierenden, die noch nicht auf der Etappe der Abschlussprüfung angekommen waren, sprich noch nicht angetreten waren und auch nicht vor hatten, beim nächsten Termin anzutreten. Sie wurden also als die Gruppe mit weniger Expertise definiert. In der zweiten Gruppe befanden sich Student_innen gegen Ende des Studiums, die alle Teile der Abschlussprüfung positiv bestanden haben. Diese Einteilung erscheint sinnvoller als eine Einteilung nach absolvierten Semestern, da die Studierenden aus Erfahrung oftmals länger als die geplante Mindeststudienzeit studieren¹². Alle teilnehmenden Personen hatten Englisch als A- oder B-Sprache und waren, bis auf eine Person¹³, im Schwerpunkt Konferenzdolmetschen.

Die Teilnehmer_innen wurden im Vorhinein darüber informiert, dass es sich um einen Versuch im Rahmen einer Masterarbeit zum Thema Qualität und dessen Bewertung im Konferenzdolmetschen handelte. Ihnen wurde erklärt, dass sie an zwei Terminen aufgezeichnete Verdolmetschungen aus dem Englischen ins Deutsche mit Hilfe eines Fragebogens bewerten sollen. Allen Personen wurde versichert, dass ihre Daten anonymisiert verwendet werden würden.

¹¹ Zwar meldeten sich insgesamt ein paar Kolleg_innen mehr, mussten jedoch ausgeschlossen werden, da sonst keine vergleichbaren Gruppen möglich gewesen wären. So wäre beispielsweise eine Gruppe von zwei Erstsemestrigen für die vorliegende Arbeit nicht groß genug gewesen.

¹² Tatsächlich gaben die Versuchspersonen an, bereits zwischen 4 und 7 Semester lang im Master zu studieren.

¹³ Die Ausnahme stellte eine Person dar, die offiziell den Schwerpunkt Dialogdolmetschen ausgewählt hat. Ein Blick ins Curriculum des ZTWs zeigt jedoch, dass auch in diesem Schwerpunkt Lehrveranstaltungen zum Simultandolmetschen absolviert werden (ZTW 2005:12f.). Zudem war die Person Teil der ersten Gruppe und war somit noch nicht zur Modulprüfung im Dialogdolmetschen angetreten.

5.1.3. Vorgehensweise und Durchführung

Da der Versuch in der Arbeit von Behr (2013a) als Ausgangspunkt dient, wurde versucht, diesem im Ablauf möglichst nahe zu kommen, wo immer es für den vorliegenden Versuch sinnvoll erschien. Behr (2013a) führte den Versuch so durch, dass die Prüfer_innen zunächst nach jeder Aufnahme eine Gesamtnote, basierend auf dem Gesamteindruck, vergeben sollten. Dem folgte die Bewertung der einzelnen Aspekte. Nach der Bewertung aller Aufnahmen konnten die Teilnehmer_innen die Noten in einem separaten Feld abändern. Erst am Ende der jeweiligen Termine wurden die Personen gebeten, einmalig den Stimmungsfragebogen auszufüllen (Behr 2013a:277f.). Diese Vorgehensweise wurde auch für den Versuch im Rahmen dieser Arbeit übernommen. Um die Schritte nochmals besser zu trennen und zu verhindern, dass Teilnehmer_innen durch die Unterlagen blättern und Spekulationen über den Stimmungsfragebogen anstellten, wurden die einzelnen Unterlagen jedoch nacheinander verteilt, anstatt in einer Mappe (Behr 2013a:277) übergeben zu werden. So wurden den Teilnehmer_innen nach Erklärung des Ablaufs mittels Anleitung die Bewertungsbögen ausgeteilt und anschließend die Verdolmetschungen abgespielt. Nach jeder Aufnahme hatten die Versuchspersonen Zeit, eine Gesamtnote zu vergeben und anschließend die Einzelkriterien zu bewerten.

Während die Gruppe der Prüfer_innen in der Studie von Behr (2013a) auch ein Transkript der Originalrede zu Verfügung hatte, um in üblicher Weise Notizen machen zu können (Behr 2013a:277), wurde den Teilnehmer_innen in der hier durchgeführten Studie diese Möglichkeit nicht geboten. Stattdessen wurde gleichzeitig zur Verdolmetschung die Originalrede eingespielt. Die teilnehmenden Personen hatten somit die Möglichkeit, beide Audioaufnahmen parallel mitzuhören. Dies ist aus Erfahrung in den Lehrveranstaltungen am ZTW üblicher, als mit Originaltexten zu arbeiten. Ziel war es, eine möglichst authentische Situation zu simulieren, in denen die Studierenden die Rolle der Zuhörer_innen einnehmen und sich Verdolmetschungen von Kolleg_innen anhören, um diese anschließend zu bewerten. Dies wird am ZTW in einigen Veranstaltungen durchaus im Rahmen von Peerfeedbacks gemacht. Dabei hören die Teilnehmer_innen in den Dolmetschlehrveranstaltungen anderen Studierenden beim Dolmetschen zu. Üblicherweise können Dank der Dolmetschanlage am Institut gleichzeitig sowohl die Verdolmetschung als auch das Original gehört werden und nach eigenen Präferenzen bezüglich Lautstärke und Seite der Kopfhörer eingestellt werden. Diese Möglichkeit wurde den Teilnehmer_innen auch im Rahmen des Versuchs geboten. Zusätzlich wurde ein leeres Blatt Papier zur Verfügung gestellt, das bei Bedarf für Notizen verwendet

werden konnte. Mit Ausnahme einer Person, die zu beiden Terminen nichts notiert hat, nutzen alle Teilnehmer_innen diese Möglichkeit für Notizen.

Nach dem Abspielen und dem Bewerten der drei Aufnahmen wurden die Versuchspersonen gebeten, einen Fragebogen mit persönlichen und studiumsspezifischen Angaben auszufüllen. Auf diesem befand sich ein Hinweis, der auf den Zusatzfragebogen zur Stimmung verwies. Dieser basierte auf dem Vermerk von Behr (2013a:341) und sollte eine unscheinbare Erklärung für die Erhebung der Stimmung bieten. Es wurde versucht, keine Aufmerksamkeit auf diesen Teil des Versuchs zu lenken und den Fragebogen unscheinbar wirken zu lassen. Beim zweiten Mal wurde dieser Fragebogen nicht mehr ausgeteilt. Stattdessen wurde in unauffälliger und beiläufiger Weise darauf verwiesen, dass die persönlichen und studiumsspezifischen Daten bereits vorliegen würden, und somit als letzter Schritt lediglich noch der Stimmungsfragebogen auszufüllen wäre. Dennoch wurde den Studierenden auch die Möglichkeit gegeben, aus ihrer Sicht relevante Kommentare zu vermerken, wie dies zum ersten Termin im Fragebogen zu den personenbezogenen Daten möglich war.

Auf die Bedeutung des Kontexts beim Dolmetschen wurde in vorangegangenen Abschnitten (Kapitel 2.3.2. bzw. 2.3.2.1.) eingegangen. Und auch wenn es sich bei dem Versuch im Rahmen dieser Arbeit um kein Experiment im eigentlichen Sinne handelt, soll an dieser Stelle erwähnt werden, dass „the interpreting process and/or product being analysed within an experiment will feature an interpretation removed from context“ (Hale & Napier 2013:171). Behr (2013a:279) sorgte dafür, dass ihr Versuch in einer weitgehend gewohnten Umgebung und möglichst ohne Ablenkungen durchgeführt wurde. Auch für den hier durchgeführten Versuch wurde darauf besonders Wert gelegt. Demnach wurde für den Versuch der Hörsaal 1 bzw. Seminarraum 5 am ZTW in Wien reserviert, die jeweils über eine Dolmetschanlage verfügen und auch für Lehrveranstaltungen genutzt werden. Das Quasi-Experiment wurde an zwei bzw. drei Terminen vor Ort mit einem Abstand von drei Wochen durchgeführt. Am ersten Termin konnten alle Teilnehmer_innen anwesend sein, während der zweite Termin geteilt wurde und an zwei aufeinander folgenden Tagen stattfand. Grund dafür war die eingeschränkte Verfügbarkeit der teilnehmenden Studierenden. Dennoch wurde genaustens darauf geachtet, alle Termine auf die gleiche Art und Weise durchzuführen und den Ablauf beizubehalten. Bedauerlicherweise konnte eine Person an keinem der letzten Termine persönlich vor Ort sein, weshalb eine alternative Lösung gefunden werden musste. Für eine Person musste der zweite

Termin einige Tage später online über die Plattform Zoom¹⁴ abgehalten werden. Dennoch wurde versucht, dem Ablauf vor Ort so nahe wie möglich zu kommen. Er wurde auch so geplant, dass die teilnehmende Person die Möglichkeit hatte, durch verschiedene Kopfhörer die Aufnahmen und das Original parallel zu hören.

Behr (2013a) präsentierte beiden Versuchsgruppen dieselben Aufnahmen, die jeweils circa 20 Minuten lang waren. In einer Fußnote merkt sie jedoch an, dass für die User_innen-Gruppe etwa halb so lange Aufnahmen besser gewesen wären (Behr 2013a:274). Für den Versuch im Rahmen dieser Arbeit wurden daher kürzere Aufnahmen erstellt, die jeweils eine Länge von etwa 10 Minuten aufwiesen. Die benötigten Dolmetschaufnahmen wurden im Vorhinein eigens für die Verwendung im Rahmen der vorliegenden Arbeit aufgezeichnet. Hierfür wurden Kolleg_innen gebeten, einen Text aus dem Englischen ins Deutsche zu dolmetschen. Bei den Personen handelte es sich um Absolvent_innen bzw. Studierende, die kurz vor dem Abschluss stehen, jedoch länger nicht mehr an Lehrveranstaltungen am Zentrum teilgenommen haben. Dadurch sollte eine möglichst hohe Wahrscheinlichkeit erreicht werden, dass die Versuchsteilnehmer_innen die Dolmetscher_innen nicht kennen bzw. erkennen, auch wenn dies nicht gänzlich ausgeschlossen werden konnte. Als Ausgangstext wurde eine Rede der US-amerikanischen Veterinärparasitologin Susan Little ausgewählt, die im Rahmen einer TEDx Veranstaltungen gehalten wurde, und frei zugänglich auf der Plattform YouTube zu finden war (YouTube o.J.). Die TEDx-Formate werden am ZTW häufig zu Übungszwecken in den Lehrveranstaltungen verwendet, weshalb davon ausgegangen werden konnte, dass sie sowohl den Dolmetscher_innen als auch den Versuchsteilnehmer_innen bekannt waren. Die Dolmetscher_innen erhielten vorab eine Information über das Thema der Rede und ein paar Schlüsselvokabel und wurden anschließend gebeten, ihre Verdolmetschung aufzuzeichnen. Somit standen für den Versuch fünf Verdolmetschungen des oben beschriebenen Videos von fünf unterschiedlichen Dolmetscher_innen zur Verfügung. Auch den Dolmetscher_innen wurde versichert, dass ihre Daten in anonymisierter Form verwendet werden würden.

Behr (2013a) bat in ihrem Versuch die Prüfer_innen-Gruppe um die Bewertung von fünf Aufnahmen beim ersten Termin, sowie von vier Aufnahmen beim zweiten Termin. Die Gruppe der Medizinstudent_innen bzw. der User_innen sollte an beiden Terminen drei Dolmetschungen bewerten (Behr 2013a:273f.). Die zu untersuchende Verdolmetschung wurde jeweils an beiden Terminen als dritte Aufnahme eingespielt. Hierfür wurde eine Dolmetschung gewählt, die keine allzu markanten Charakteristika aufwies, wie besonders auffällige Versprecher oder Fehler. So

¹⁴ Dies entspricht zwar nicht genau dem Setting, wie es momentan am Zentrum üblich ist, dennoch ist das Online-Format den meisten Studierenden aus der Zeit der Lockdowns bekannt.

sollte verhindert werden, dass die Teilnehmenden sich an diese beim letzten Termin erinnerten (Behr 2013a:279). Auch im Rahmen der vorliegenden Arbeit wurde die zu untersuchende Aufnahme, die an beiden Terminen eingespielt werden sollte, aus den fünf verfügbaren Aufzeichnungen sorgfältig ausgesucht. Es wurde darauf geachtet, dass diese keine besonders herausstechenden Eigenschaften aufwies, um die Wahrscheinlichkeit einer Wiedererkennung so weit wie möglich zu minimieren¹⁵. Wie Behr (2013a) anmerkte, war in Hinblick auf die Reihenfolge der Aufnahmen darauf zu achten, dass die zu untersuchende Verdolmetschung so eingespielt wird, dass an beiden Terminen zuvor dieselbe Anzahl an Dolmetschungen präsentiert wurde. So kann von einem gleichen Maß in Hinblick auf die Gewöhnung, die Ermüdung oder eine nachlassende Konzentration ausgegangen werden (Behr 2013a:279). Bei jedem Termin wurden also insgesamt drei Dolmetschungen des oben beschriebenen TEDx-Videos für die Bewertung eingespielt. Die gewählte, zu untersuchende Aufnahme wurde dabei an allen Terminen als zweite Dolmetschung präsentiert.

5.2. Methoden der Datenerhebung

Für die Erhebung der für das Forschungsinteresse relevanten Daten wurden zwei separate Fragebögen eingesetzt, anhand derer die Bewertungen der Aufnahmen und die Stimmungen der Versuchsteilnehmer_innen erhoben wurden. Diese wurden, wie oben beschrieben (siehe Kapitel 5.1.3.), von den Teilnehmer_innen im Rahmen des empirischen Versuchs ausgefüllt. Die Konzeption dieser Bögen orientierte sich an jenen von Behr (2013a).

An dieser Stelle sei angemerkt, dass Behr (2013a) sowohl für den Bewertungs- als auch Stimmungsfragebogen eine visuelle Analogskale (VAS) verwendete (Behr 2013a:281). Solche Skalen sind horizontale Linien ohne Markierungen und weisen zumeist eine Länge von 100mm auf und werden an den jeweiligen Enden mit verbalen Ankern versehen. Beim Beantworten wird eine Markierung auf der Linie gesetzt. Visuelle Analogskalen werden zumeist im medizinischen Bereich angewendet, beispielsweise um Schmerz zu messen (Funke 2016:244). Mit einer solchen Skala ist die Anzahl der möglichen Antwortkategorien nicht beschränkt (Funke 2006:4827). Doch „[w]enn digitale Datenerhebungsmöglichkeiten nicht angewandt werden können, fallen die Nachteile des manuellen Ausmessens der Visuellen Analogskalen

¹⁵ Dennoch konnte nicht verhindert werden, dass eine Versuchsperson die Aufnahme beim zweiten Termin wiedererkannte und dies schriftlich in einem Kommentar festhielt.

erheblich ins Gewicht“ (Funke 2006:4827). Zudem merkt Behr (2013a) in einer Fußnote an, dass die Versuchspersonen in ihrer Studie teilweise äußerten, dass ihnen eine Orientierung ohne Einteilung auf den Skalen schwer fiel (Behr 2013a:283). Daher wurden für die Fragebögen im Rahmen dieses Versuchs andere Skalen gewählt. Im Folgenden soll näher auf die konzipierten Fragebögen eingegangen werden.

5.2.1. Fragebogen zur Bewertung der Verdolmetschungen

Behr (2013a) konzipierte ihren Bewertungsbogen anhand von sechs Kriterien. Diese umfassten einerseits die inhaltsbezogenen Parameter „Vollständigkeit“, „Grammatik“ und „Richtigkeit“, und andererseits die weniger objektiv überprüfbaren Einzelaspekte „Funktionalität“, „Prosodie“ und „Ausdruck“ (Behr 2013a:303). Den bereits oben erläuterten Anmerkungen von Behr folgend, dass beim Bewerten in der Praxis eine eher geringere Kriterienanzahl zum Einsatz kommt (Behr 2013b:87), sollte auch für den Fragebogen in dieser Arbeit eine geringere Anzahl beibehalten werden. Ausgehend von Behrs (2013a) Fragebogen sowie den Beurteilungskriterien, die am ZTW bei den Modulprüfungen zum Simultandolmetschen zum Einsatz kommen¹⁶, wurde ein neuer Fragebogen konzipiert. Einige Kriterien von Behrs (2013a) Fragebogen wurden durch Bezeichnungen ersetzt, die (eher) den Kriterien im Bewertungsbogen des ZTWs entsprachen.

Insgesamt umfasste der Bewertungsbogen für den empirischen Versuch dieser Arbeit sieben Qualitätskriterien, die in inhaltliche und kommunikativ-sprachliche Aspekte unterteilt werden können (siehe Tabelle 2). Die Kriterienübersicht des ZTWs teilt die Aspekte jedoch in drei Blöcke ein – „Inhalt“, „Zielsprache“, sowie „Dolmetsch-Fertigkeiten & Präsentation“ (ZTW o.J.). Um diese vertraute Unterteilung beizubehalten, wurden diese drei Blöcke übernommen und für den letzten Block der Parameter „Flüssigkeit“ auch in den Fragebogen aufgenommen (siehe Anhang). Um die Notenvergabe zu erheben, wurden Ordinalskalen angewendet.

¹⁶ Aus Erfahrung kann gesagt werden, dass auf die Kriterien auch in den Lehrveranstaltungen Bezug genommen wird und sie teilweise auch als Grundlage für Semesternoten herangezogen werden. Daher war anzunehmen, dass die Studierenden am ZTW mit ihnen vertraut sind.

Tabelle 2: Inhaltliche und kommunikativ-sprachliche Qualitätskriterien im Bewertungsbogen, eigene Aufstellung, basierend auf Behr (2013a) und ZTW (o.J.)

inhaltliche Qualitätskriterien	kommunikativ-sprachliche Qualitätskriterien
- Vollständigkeit - Inhaltliche Korrektheit - Syntax, Grammatik	- Angemessenheit fürs Zielpublikum - Lexik (lexikalische/stilistische Aspekte) - Prosodie - Flüssigkeit

5.2.2. Stimmungsfragebogen

Eine einfache und schnelle Messung der Stimmung kann beispielsweise durch das in Kapitel 4.1.2. beschriebene Affektgitter nach Mendelsohn et al. (1989:494) erfolgen. Dabei handelt es sich um eine Skala „potentially suitable for any study that requires judgments about affect of either a descriptive or a subjective kind“ (Mendelsohn et al. 1989:493). Dennoch ist das Affektgitter nicht für alle Forschungszwecke geeignet und manchmal erscheint es sinnvoll, stattdessen einen Fragebogen zu wählen (Mendelsohn et al. 1989:499). Denn „[w]hen the researcher’s focus is on the individual subject and when time is available, a multiple-item questionnaire should be used if possible“ (Mendelsohn et al. 1989:499). Das Affektgitter setzt nämlich auch eine Auseinandersetzung mit längeren Anweisungen und dem Gitter selbst voraus. Bei einmaliger Anwendung beispielsweise könne man auch keine Zeitersparnis erreichen (Mendelsohn et al. 1989:499). Aus diesen Gründen entschied sich auch Behr (2013a:288) in ihrer Studie, das Gitter selbst nicht heranzuziehen, es jedoch als Ausgangspunkt für einen eigens angefertigten Stimmungsfragebogen zu nutzen, „da es sehr umfassend diejenigen Begriffe bzw. Stimmungen abfragt, die für das generelle Befinden einer Person als relevant und erschöpfend gelten“ (Behr 2013a:288). So wurde die Stimmung anhand der zwei Dimensionen Gestimmtheit und Aktivierung erhoben. Ihr Stimmungsfragebogen bestand weiters aus zwei Teilen, die Informationen zur Stimmung der Versuchsteilnehmer_innen jeweils am Tag des Versuchs sowie für die fünf Tage zuvor erheben sollte (Behr 2013a:289f.). Sie vermutete, „dass anhand der Vergleichsmöglichkeit dieser beiden Angaben stärker gewährleistet ist, tatsächlich die generelle Stimmung zu erfassen und eben nicht die, die in Folge der Bewertung punktuell verspürt wurde“ (Behr 2013a:290). Sie übertrug daher die Begriffe des *Affect Grids* zunächst ins Deutsche, wobei für die zwei Teile des Stimmungsfragebogens jeweils verschiedene Begriffe festgelegt wurden. Diese sollten so

verständlich und so wenig ambivalent wie möglich sein (Behr 2013a:289). Diese Übertragung bei Behr (2013a) ist in Tabelle 3 zu sehen.

Die Auswertung bei Behr (2013a) der erhobenen Stimmung erfolgte dann anhand eines Rasters, in dem die Durchschnittswerte für die zwei Stimmungsdimensionen berechnet werden konnten, die gemeinsam schließlich den Stimmungswert erhaben (Behr 2013a:290f.). Dieser ist in Tabelle 4 dargestellt.

Tabelle 3: Das Übertragen der Items aus dem Affektgitter von Mendelsohn et. al (1989) nach Martina Behr (2013a:289)

Affect Grid	Fragebogen Teil 1	Fragebogen Teil 2
arousal – sleepiness	voller Energie – müde	hellwach – träge
excitement – depression	beschwingt – niedergeschlagen	neugierig – gelangweilt
pleasure – misery	gut gelaunt – schlecht gelaunt	wohl – unwohl
contentment - stress	erholt – gestresst	ausgeglichen – angespannt

Tabelle 4: Raster zur Auswertung der Stimmung (Gestimmtheit und Aktivität) nach Behr (2013a:291)

Vpn	Gestimmtheit: positiv / negativ		Aktivität: aktiv / Passiv	
	erholt / gestresst	Angaben „letzten 5 Tage“	voll Energie / müde	
	gut/schlecht gelaunt		beschwingt / niedergeschlagen	
Ø:				Ø:
	ausgeglichen / angespannt	Angaben „heute“	hellwach / träge	
	wohl / unwohl		neugierig / gelangweilt	
Ø:				Ø:
Gesamtwert Gestimmtheit:			Gesamtwert Aktivierung:	

Der Stimmungsfragebogen, der für die vorliegende Arbeit konzipiert wurde, basiert auf dem soeben beschriebenen Bogen nach Behr (2013a). Dabei wurde zum einen die Unterteilung nach der Stimmung der letzten fünf Tage und der Stimmung am gegebenen Tag übernommen. Weiters wurden auch die von Behr (2013a) übertragenen Begrifflichkeiten beibehalten. Wie bereits oben angedeutet, wurden statt virtuellen Analogskalen jedoch eine Ratingskala gewählt, die an beiden Enden mit den verbalen Ankern versehen wurde (siehe Anhang). Für die Auswertung der erhobenen Stimmung wurde wiederum der Auswertungsraster nach Behr (2013a:291) verwendet.

5.3. Ergebnisse

Im Folgenden werden die Auswertungen, die anhand der im empirischen Versuch erhobenen Daten durchgeführt wurden, präsentiert. Entsprechend der in Kapitel 5.1.1. beschriebenen Forschungsfragen und Ausgangspunkte wurde an einigen Stellen zwischen Studierenden vor der Abschlussprüfung und jenen, die sich am Ende des Studiums befinden, unterschieden. Gruppe 1 (Vpn 1 bis 4) umfasst somit Studierende mit vergleichsweise geringerem Studienfortschritt. Gruppe 2 (Vpn 5 bis 8) bezieht sich im Folgenden auf Student_innen mit mehr Expertise, die bereits die abschließende Modulprüfung bestanden haben.

5.3.1. Gesamtbenotung derselben Dolmetschung zu zwei verschiedenen Zeitpunkten

Die Betrachtung der vergebenen Gesamtbeurteilung hat gezeigt, dass die Mehrheit aller Teilnehmer_innen beim Bewerten der Verdolmetschung an den zwei Terminen jeweils unterschiedliche Noten vergeben haben. Abbildung 2 zeigt die Differenz der vergebenen Gesamtbenotung für beide Gruppen. Nur zwei Personen (Vpn 4 aus Gruppe 1 und Vpn 7 aus Gruppe 2) bewerteten die Aufnahme gleich. Die anderen Versuchspersonen vergaben Gesamtnoten, die sowohl um einen Grad schlechter als auch besser waren. Eine Person aus Gruppe 1 vergab beim zweiten Termin eine um zwei Grad schlechtere Note. Insgesamt betrachtet wurde beim zweiten Termin dreimal eine vergleichsweise bessere, und dreimal eine vergleichsweise schlechtere Note für dieselbe Dolmetschung vergeben.

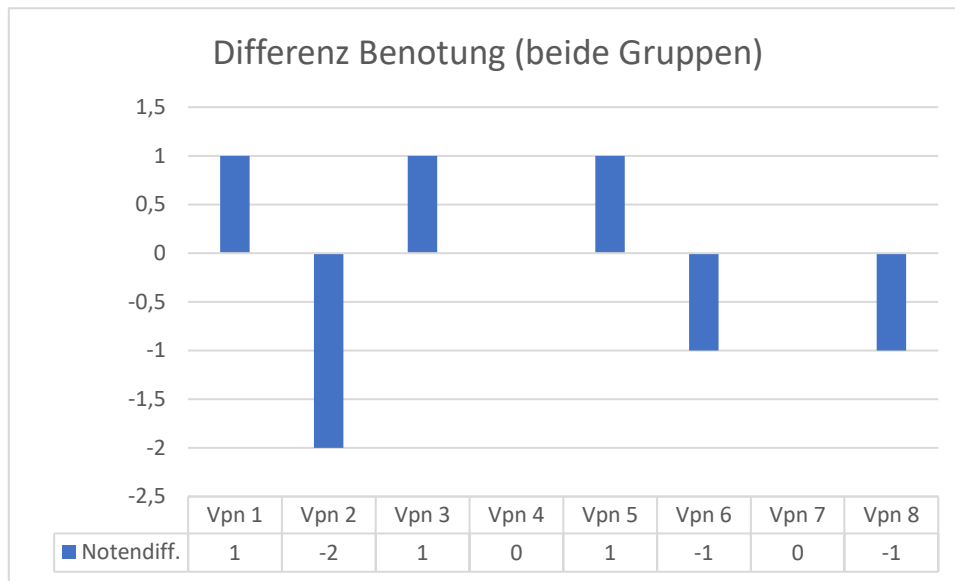


Abbildung 2: Differenz bei der Benotung zwischen erstem und zweitem Termin - beide Gruppen (eigene Erhebung)

Die Annahme, dass es zu Unterschieden zwischen den Versuchspersonen in Hinblick auf die Notengebung kommen würde, hat sich in diesem Versuch somit allgemein bestätigt. Auch die angenommen Bewertungsunterschiede bei den einzelnen Personen zeigten sich bei sechs von acht Personen und konnten so allgemein bestätigt werden. Jedoch kann an dieser Stelle noch kein klar ersichtlicher Unterschied in Hinblick auf Expertise getroffen werden. Denn auch wenn die größte Notendifferenz in Gruppe 1 gefunden wurde, schwanken die Noten in beiden Gruppen. Zudem sind, wie erwähnt, in beiden Gruppen Personen zu finden, die die gleiche Note vergeben haben.

Wie in Tabelle 5 ersichtlich, wurden an beiden Terminen für die untersuchte Aufnahme Noten von 1 (Sehr Gut) bis 4 (Genügend) vergeben, wobei die Note 2 (Gut) am häufigsten vergeben wurde. Die Note 4 wurde dabei beide Male in Gruppe 1 vergeben. Somit lag das Notenspektrum in Gruppe 1 beide Male bei 2 (Gut) bis 4 (Genügend). Im Vergleich dazu vergab Gruppe 2 beim ersten Mal nur Noten von 1 (Sehr Gut) bis 2 (Gut). Beim zweiten Mal wurde auch einmal die Note 3 vergeben. Die Durchschnittsnote insgesamt betrug 2,25 beim ersten und 2,38 beim zweiten Termin, was gerundet somit eine Note von 2 (Gut) ergibt. Betrachtet man die Notenwerte getrennt nach den beiden Gruppen, ist zu erkennen, dass der Mittelwert der vergebenen Noten bei Gruppe 1 an beiden Terminen eine Durchschnittsnote von 2,75 ergab, während Gruppe 2 die Leistung beim ersten Termin durchschnittlich mit 1,75 und beim zweiten Termin im Durchschnitt mit 2 bewertet hat. Gerundet entspricht dies einer Bewertung mit 3 (Befriedigend) durch Gruppe 1 an beiden Terminen, und der Note 2 (Gut) durch Gruppe 2, ebenfalls an beide Terminen. Somit kann auch festgehalten werden, dass in diesem Versuch

beobachtet wurde, dass die einzelnen Personen zwar in ihrer Bewertung schwanken, aber alle Studierenden gemeinsam die Verdolmetschung an beiden Terminen durchschnittlich gleich evaluierten. Dies entspricht den Beobachtungen bei Behr (2013a), in deren Versuch zwar nicht die Personen einzeln, aber die Dolmetscher_innen bzw. Prüfer_innen als Gruppe der gleichen Aufnahme an den zwei Versuchsterminen die gleiche Note gegeben haben (Behr 2013a:319).

Tabelle 5: Absolute Häufigkeiten und Mittelwerte der vergebenen Noten beider Gruppen an beiden Terminen (eigene Erhebung)

Termin 1	Absolute Häufigkeit Gruppe 1	Absolute Häufigkeit Gruppe 2	Absolute Häufigkeit (gesamt)
„Sehr Gut“	0	1	1
„Gut“	2	3	5
„Befriedigend“	1	0	1
„Genügend“	1	0	1
„Nicht Genügend“	0	0	0
Notendurchschnitt	2,75	1,75	2,25
Termin 2	Absolute Häufigkeit Gruppe 1	Absolute Häufigkeit Gruppe 2	Absolute Häufigkeit (gesamt)
„Sehr Gut“	0	1	1
„Gut“	2	2	4
„Befriedigend“	1	1	2
„Genügend“	1	0	1
„Nicht Genügend“	0	0	0
Notendurchschnitt	2,75	2	2,38

Am Ende des Versuchs, nachdem alle Aufnahmen evaluiert worden waren, wurde den Versuchspersonen angeboten, die von ihnen vergebene Gesamtbeurteilung abzuändern. Beim ersten Termin wurde diese Möglichkeit von einer Person in Anspruch genommen, beim zweiten Termin änderten zwei Personen ihre Gesamtbeurteilung. Alle drei Änderungen bezogen sich jedoch nicht auf die untersuchte Aufnahme Nr. 2, sondern wurden alle für Aufnahme Nr. 1 durchgeführt. Es handelte sich dabei sowohl um Abänderungen auf bessere als auch schlechtere Noten.

5.3.2. Zusammenhang zwischen Gesamtnotengebung und Stimmung

Für die Berechnung und Auswertung der Stimmung wurden die im Fragebogen verwendeten Skalen (0 bis 10) so übertragen, dass die negativen Ankerpunkte mit -5 und die positiven Adjektive mit +5 versehen wurden. Der Wert 0 (ursprünglich Wert 5 auf dem Fragebogen)

entspricht somit einem neutralen Stimmungswert. So waren letztendlich Stimmungswerte von -5 bis +5 möglich.

Die Auswertung der Stimmung hat gezeigt, dass alle Teilnehmer_innen zu den beiden Zeitpunkten unterschiedlich gestimmt waren, wobei sich die Stimmung bei zwei Personen vergleichsweise verbessert hat, während sich bei den anderen Studierenden die Stimmung verschlechtert hat. Abbildung 3 zeigt die Differenz der Stimmungswerte für beide Gruppen.

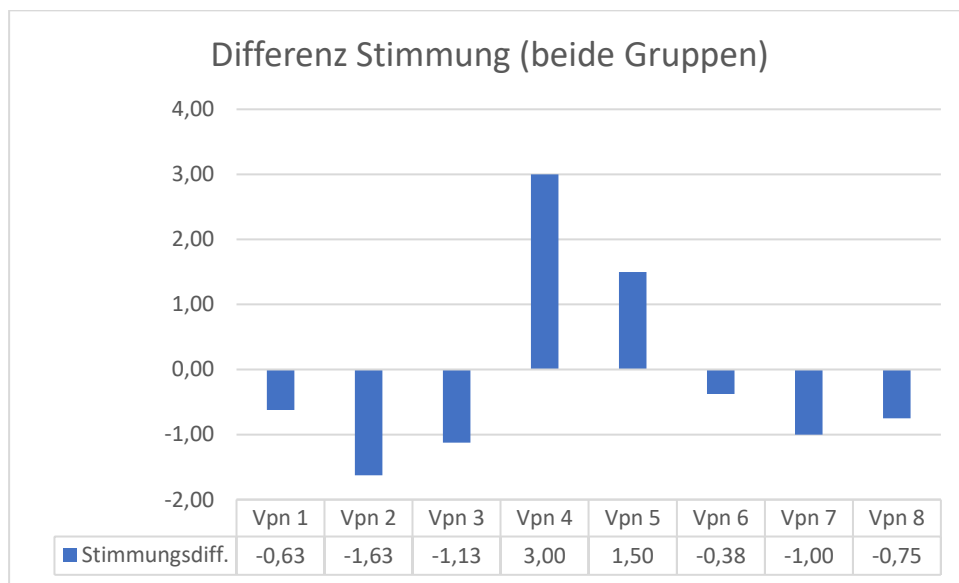


Abbildung 3: Stimmungsdifferenz beider Gruppen (eigene Erhebung)

Betrachtet man die Stimmungswerte zu den beiden Zeitpunkten im Vergleich (siehe Abbildung 4), so ist zu erkennen, dass in beiden Gruppen an beiden Terminen Personen mit negativer und positiver Stimmung waren.

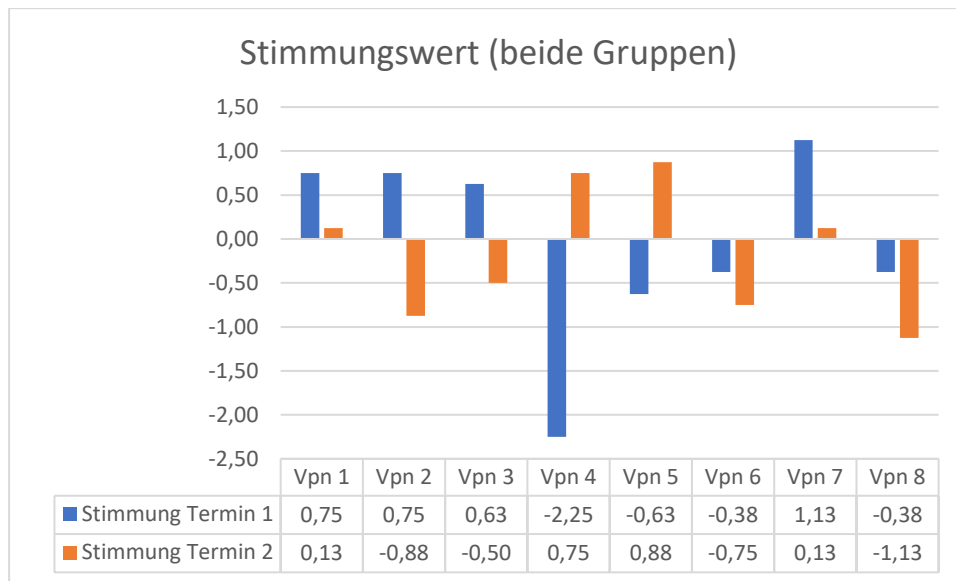


Abbildung 4: Stimmungswerte beider Gruppen an beiden Terminen (eigene Erhebung)

Dennoch muss an dieser Stelle festgehalten werden, dass die Mehrheit der Stimmungswerte um den Wert null bzw. zwischen -1,13 und +1,13 liegt. Wenn man bedenkt, dass für die Stimmung ein Wert zwischen -5 und +5 möglich war, tendieren die Werte eher zu einer neutralen Stimmung. Der ausgeprägteste (negative) Werte war -2,25 (Vpn 4 beim ersten Termin). Auch wenn die Stimmungen insgesamt nicht allzu ausgeprägt waren, so waren unter den Studierenden an beiden Terminen dennoch die eine Hälfte (eher) positiv gestimmt und die andere Hälfte (eher) negativ gestimmt. Ob sich an dieser Stelle ein Zusammenhang zwischen der Notenvergabe und der erhobenen Stimmung feststellen lassen kann, sollte anhand einer Gegenüberstellung der Noten- und Stimmungsdifferenz herausgefunden werden. Abbildung 5 zeigt diese Gegenüberstellung für Gruppe 1.

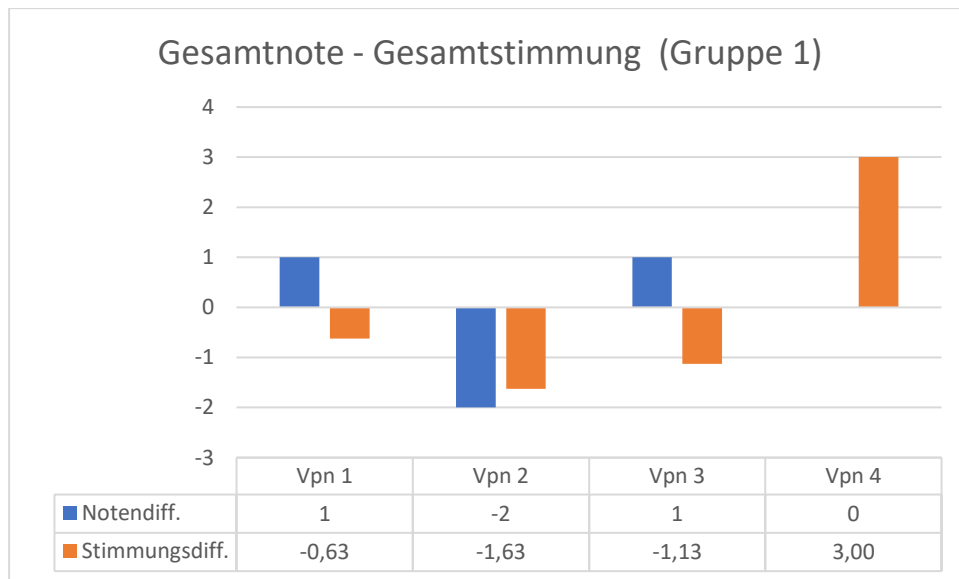


Abbildung 5: Gegenüberstellung der Noten- und Stimmungsdifferenz für Gruppe 1 (eigene Erhebung)

Betrachtet man lediglich diese Gruppe (Gruppe 1), so ist keine klare Tendenz zu erkennen. Nur bei einer Person ging die Vergabe einer schlechteren Note mit einer Verschlechterung der Stimmung einher (Vpn 2). Bei Vpn 1 und 3 war die Situation genau umgekehrt. Überraschenderweise vergab die Person mit dem größten Stimmungsunterschied (Vpn 4 mit einer Stimmungsdifferenz von +3) an beiden Terminen die gleiche Note. Die Betrachtung der zweiten, erfahreneren Gruppe zeigt ein etwas anderes Bild (siehe Abbildung 6).

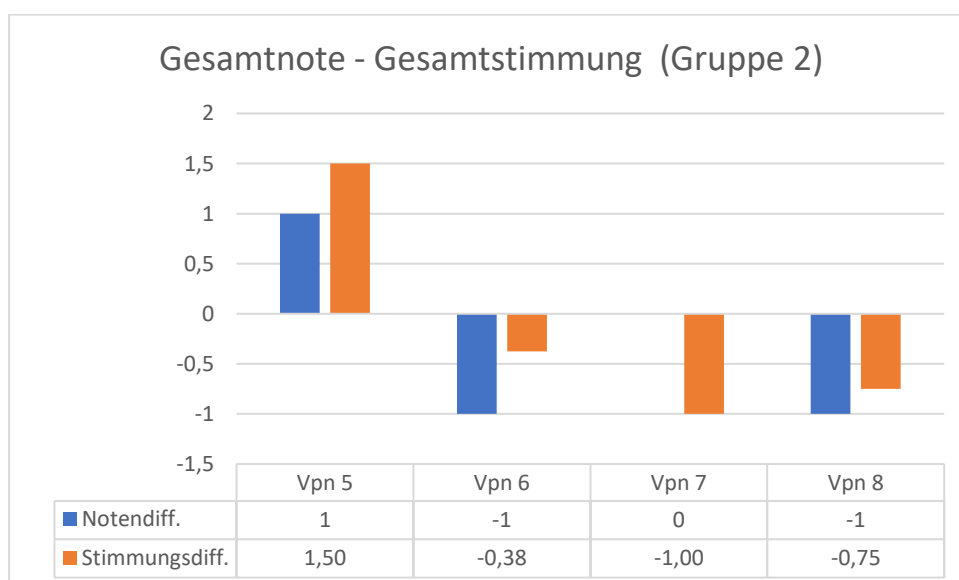


Abbildung 6: Gegenüberstellung der Noten- und Stimmungsdifferenz für Gruppe 2 (eigene Erhebung)

In dieser Gruppe ging die Verbesserung bzw. Verschlechterung der Stimmung eher mit der Verbesserung bzw. Verschlechterung der vergebenen Noten einher. Die Person mit der zweitgrößten Stimmungsdifferenz (Vpn 7) vergab jedoch beide Male dieselbe Gesamtnote. Bei den übrigen drei Personen scheint die Bewertung eher mit der Stimmung zu korrelieren¹⁷. Dies ist überraschend, da es sich laut vorgenommener Definition hierbei um die Gruppe mit mehr Expertise handelt, der laut der FAIT andere Informationen zur Bewertung zur Verfügung stehen, auf die zurückgegriffen werden kann. In der hier durchgeführten empirischen Untersuchung unter den beschriebenen Bedingungen konnte bei den erfahrenen Studierenden somit eher eine Tendenz erkannt werden. Insgesamt sind die Daten jedoch nicht ausreichend, um allgemeine Schlüsse aufzustellen. Dennoch sollte an dieser Stelle auch die getrennte Betrachtung der zwei Stimmungsdimensionen erfolgen, wie dies auch bei Behr (2013a:299ff.) vorgenommen wurde. So sollte überprüft werden, ob in diesem konkreten Versuch eine der zwei Dimensionen eher mit der Vergabe der Gesamtnote korrelierte. Dies soll zunächst für die Gruppe mit mehr Expertise erfolgen. Abbildung 7 zeigt die Gegenüberstellung der Notendifferenz und der Stimmungsdifferenz für die Dimension Gestimmtheit für Gruppe 2. Abbildung 8 zeigt diese Gegenüberstellung bei Gruppe 2 für die Dimension Aktivität.

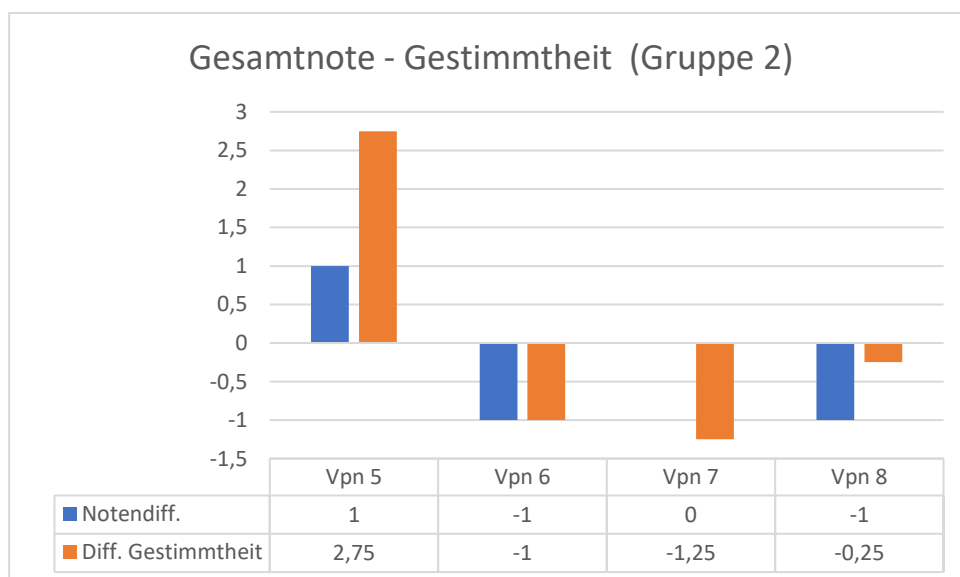


Abbildung 7: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Gestimmtheit für Gruppe 2 (eigene Erhebung)

¹⁷ An dieser Stelle sei kurz angemerkt, dass die Berechnung für Gruppe 2 einen Korrelationskoeffizienten zwischen Noten- und Stimmungsdifferenz von rund 0,76 ergab, was auf eine eher starke positive Korrelation hindeutet. Aufgrund der geringen Personenanzahl kann dies jedoch nicht als aussagekräftiges statistisches Ergebnis interpretiert werden.

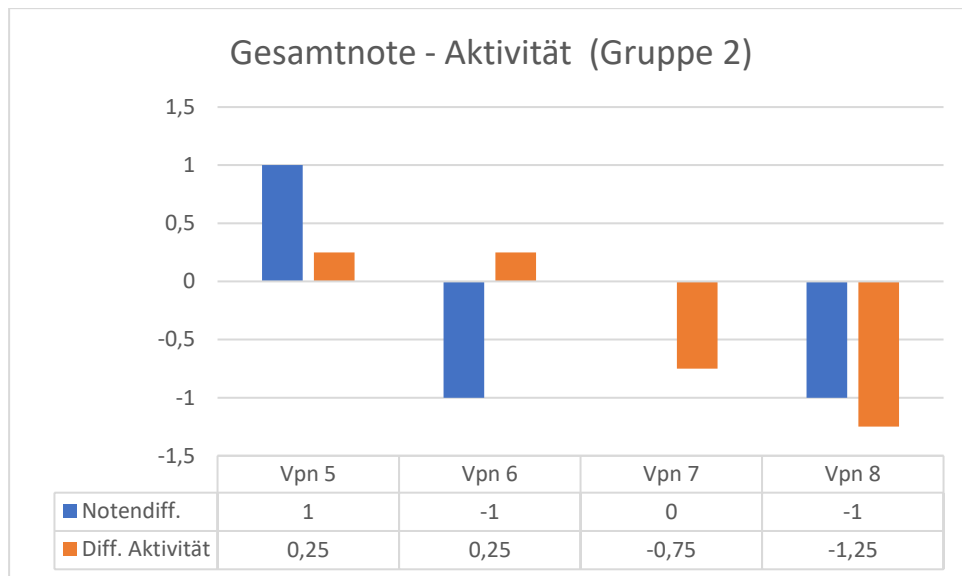


Abbildung 8: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Aktivität für Gruppe 2 (eigene Erhebung)

Dabei ist eindeutig zu erkennen, dass in dieser Untersuchung eher in der Dimension Gestimmtheit eine Tendenz zu erkennen ist. An dieser Stelle wären weitere Versuche notwendig, um zu untersuchen, ob die Dimension der Gestimmtheit im Bereich der Evaluation von Verdolmetschungen möglicherweise mehr ins Gewicht fällt und mit der Beurteilung korreliert. Eine solche nach den Gefühlsdimensionen differenzierte Betrachtung ergab bei der ersten Gruppe kein klares Bild. Im Gegensatz zur erfahreneren Gruppe war hier weder in Hinblick auf die Differenz bei der Gestimmtheit (Abbildung 9) noch bei jener der Aktivität (Abbildung 10) ein Muster zu erkennen. So schien bei dieser Gruppe keine der beiden Dimensionen die Gesamtevaluation in eine bestimmte Richtung zu beeinflussen.

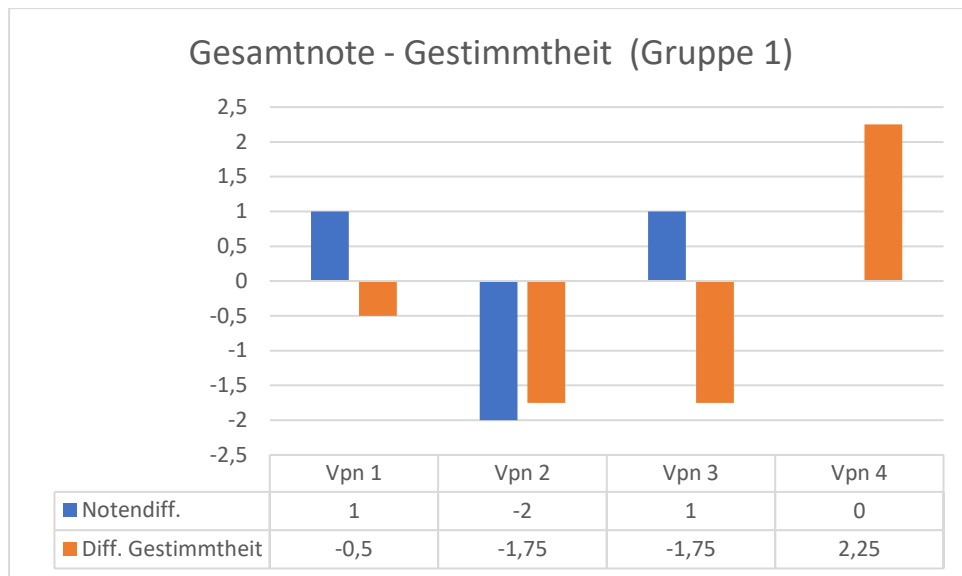


Abbildung 9: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Gestimmtheit für Gruppe 1 (eigene Erhebung)

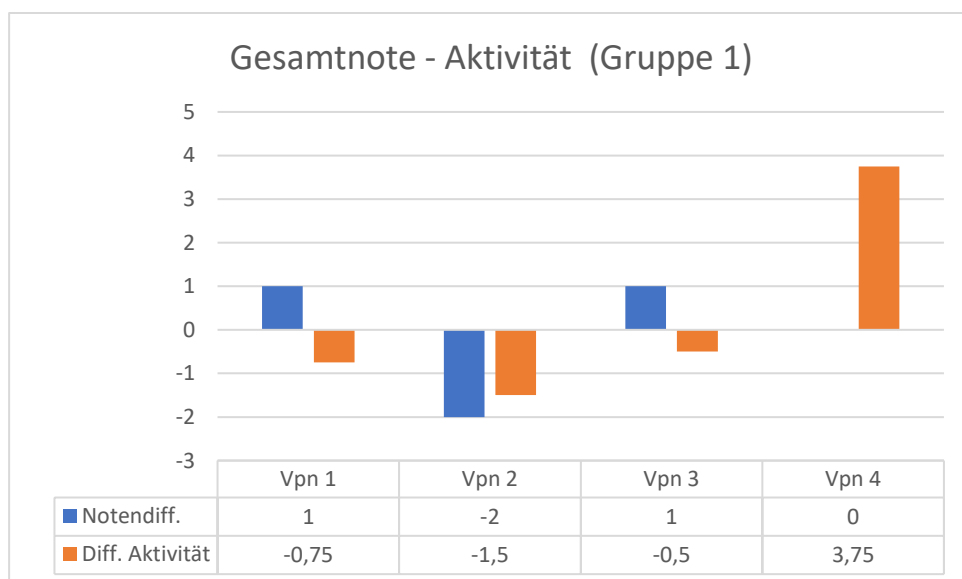


Abbildung 10: Gegenüberstellung der Notendifferenz und der Differenz bei der Stimmungsdimension Aktivität für Gruppe 1 (eigene Erhebung)

Zu den Aspekten Stimmungseinfluss und Expertise kann an dieser Stelle festgehalten werden, dass im empirischen Versuch der vorliegenden Arbeit die Unterschiede zwischen den zwei Gruppen nicht den Erwartungen gemäß der FAIT entsprachen. So war nicht zu erkennen, dass die Stimmungsunterschiede in der ersten Gruppe die Beurteilung tendenziell in eine Richtung beeinflusst hätten. Ganz im Gegenteil waren in der ersten Gruppe die zwei Extreme vertreten. Vpn 2 wies, wie bereits erläutert, den größten Notenunterschied auf und hatte zudem den

zweitgrößten Stimmungsunterschied. Vpn 4 wies den größten Stimmungsunterschied auf, vergab jedoch beide Male eine einheitliche Beurteilung. Somit kann kein Schluss gezogen werden, der besagt, dass die weniger erfahrene Gruppe eindeutig mehr auf die Stimmung zurückgreift, welche die Bewertung dann in eine bestimmte Richtung beeinflusst. Eine Tendenz war im Gegensatz dazu, wie oben beschrieben, eher in Gruppe 2 zu erkennen.

So sollte hier in einem nächsten Schritt die Noten- und Stimmungsdifferenz für beide Gruppen gemeinsam gegenübergestellt werden, um zu sehen, ob möglicherweise eine Tendenz erkennbar ist, die nicht von der unterschiedlichen Expertise dieser beiden Gruppen abhängig ist. Abbildung 11 zeigt diese Gegenüberstellung.

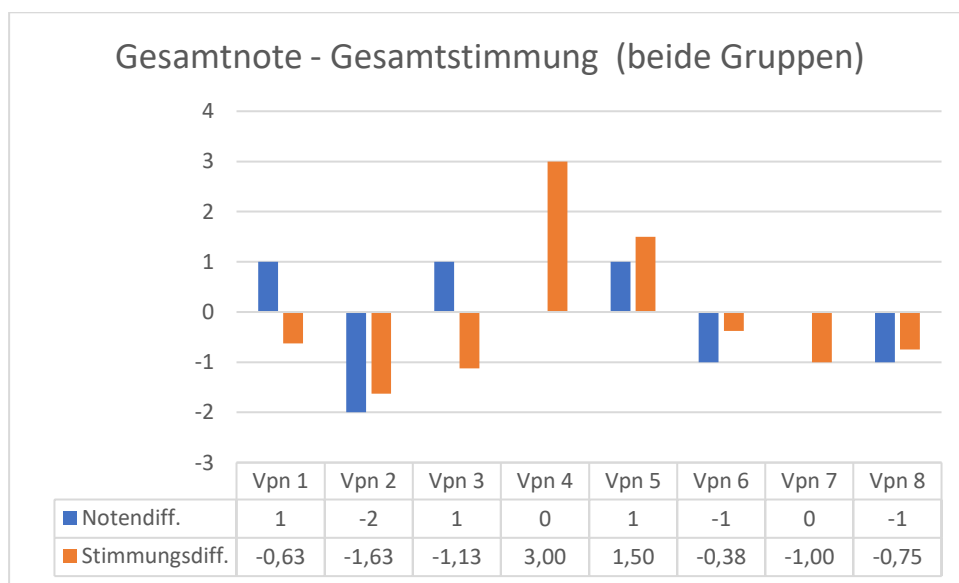


Abbildung 11: Gegenüberstellung der Noten- und Stimmungsdifferenz für beide Gruppen (eigene Erhebung)

Doch auch wenn man alle Personen gemeinsam betrachtet, ist keine eindeutige Tendenz festzustellen. Somit wird festgehalten, dass alle Personen zu den unterschiedlichen Versuchsterminen (zumindest leicht) anderer Stimmung waren, jedoch im Großen und Ganzen keine einheitliche Tendenz zu erkennen ist.

Ausgehend von den Auswertungen bei Behr (2013a) wurde schließlich auch versucht, die Personen mit den größten Stimmungsdifferenzen zwischen den zwei Terminen zusammen zu betrachten. Abbildung 12 zeigt jeweils die zwei Teilnehmer_innen mit der größten Stimmungsdifferenz in negative bzw. positive Richtung, unabhängig von ihrer Einteilung nach Expertise.

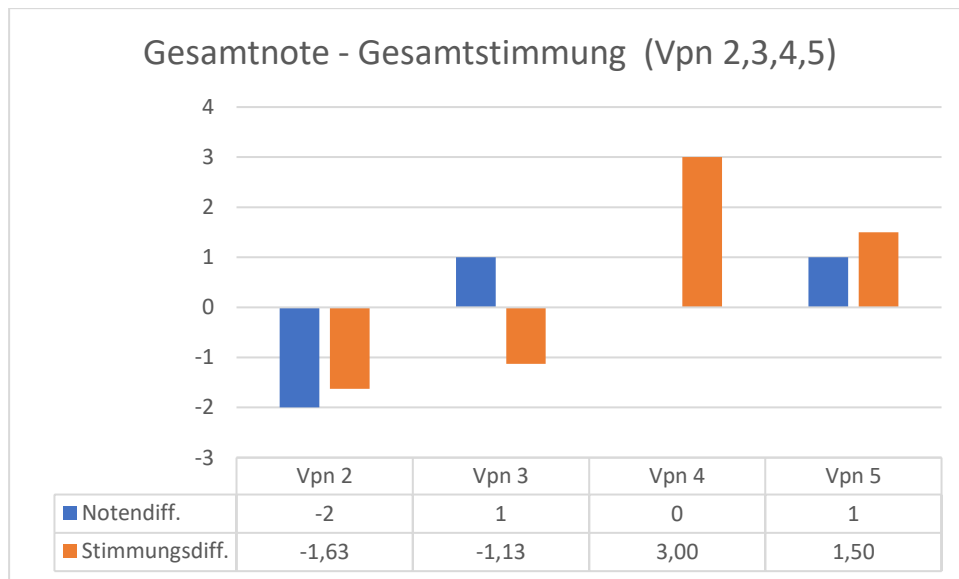


Abbildung 12: Gegenüberstellung der Noten- und Stimmungsdifferenz Vpn 2, 3, 4 und 5 (eigene Erhebung)

Anhand dieser Betrachtung wird nochmals deutlich, dass im empirischen Versuch der vorliegenden Arbeit keine aussagekräftige Korrelation zwischen der Stimmung und der Bewertung durch Studienkolleg_innen nachgewiesen wurde.

5.3.3. Bewertung der Einzelaspekte

Um der Frage nachzugehen, inwieweit sich die Bewertungen der beiden Gruppen in Hinblick auf die verschiedenen Parameter der Qualitätsbewertung unterscheiden, sollen zunächst die Notendifferenzen bei den Einzelaspekten betrachtet werden. Abbildung 13 zeigt dies für die Studierenden der Gruppe 1.

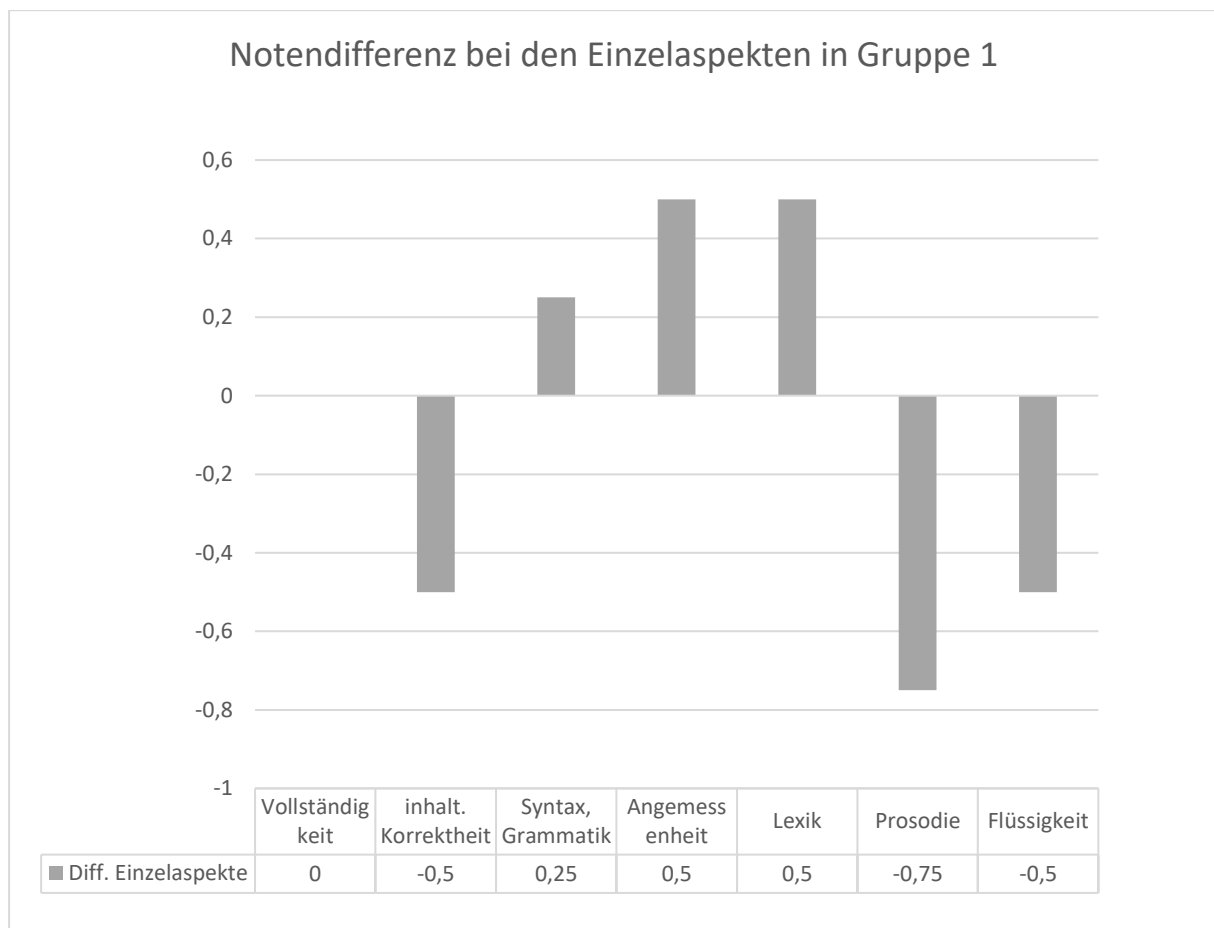


Abbildung 13: Differenz der Noten für die Einzelaspekte bei Gruppe 1 (eigene Erhebung)

Es ist zu erkennen, dass die kommunikativ-sprachlichen Aspekte stärkeren Schwankungen unterliegen. So weisen sowohl der Aspekt „Angemessenheit fürs Zielpublikum“ als auch die beiden Kriterien „Lexik“ und „Flüssigkeit“ in Gruppe 1 zwischen den Terminen eine Differenz von einem halben Notengrad auf. Der Einzelaspekt der „Prosodie“ weist mit -0,75 die größte Differenz auf. Im Gegensatz dazu ist zu erkennen, dass es bei „Syntax, Grammatik“ lediglich zu einer Differenz von 0,25 gekommen ist, während für die „Vollständigkeit“ in dieser Gruppe durchschnittlich an beiden Tagen die gleiche Note vergeben wurde. Dennoch schwank bei den inhaltsbezogenen Aspekten bei dieser Gruppe auch die inhaltliche Korrektheit um einen halben Notengrad.

In der zweiten Gruppe schwankten vor allem bei zwei der kommunikativ-sprachlichen Aspekte die Noten besonders stark. In Abbildung 14 ist zu erkennen, dass die „Prosodie“ bei den erfahreneren Studenten eine Differenz von einer ganzen Note zwischen den Terminen aufweist, gefolgt von der „Lexik“ mit einer Notendifferenz von +0,5. Die anderen zwei Aspekte in dieser Kategorie schwanken jedoch weniger stark (beide um einen viertel Notengrad). In

Hinblick auf die inhaltsbezogenen Aspekte ist zu erkennen, dass die „Syntax, Grammatik“ von Gruppe 2 an beiden Terminen durchschnittlich die gleiche Note erhalten hat, während „Vollständigkeit“ mit einer Differenz von 0,5 stärker schwankte.

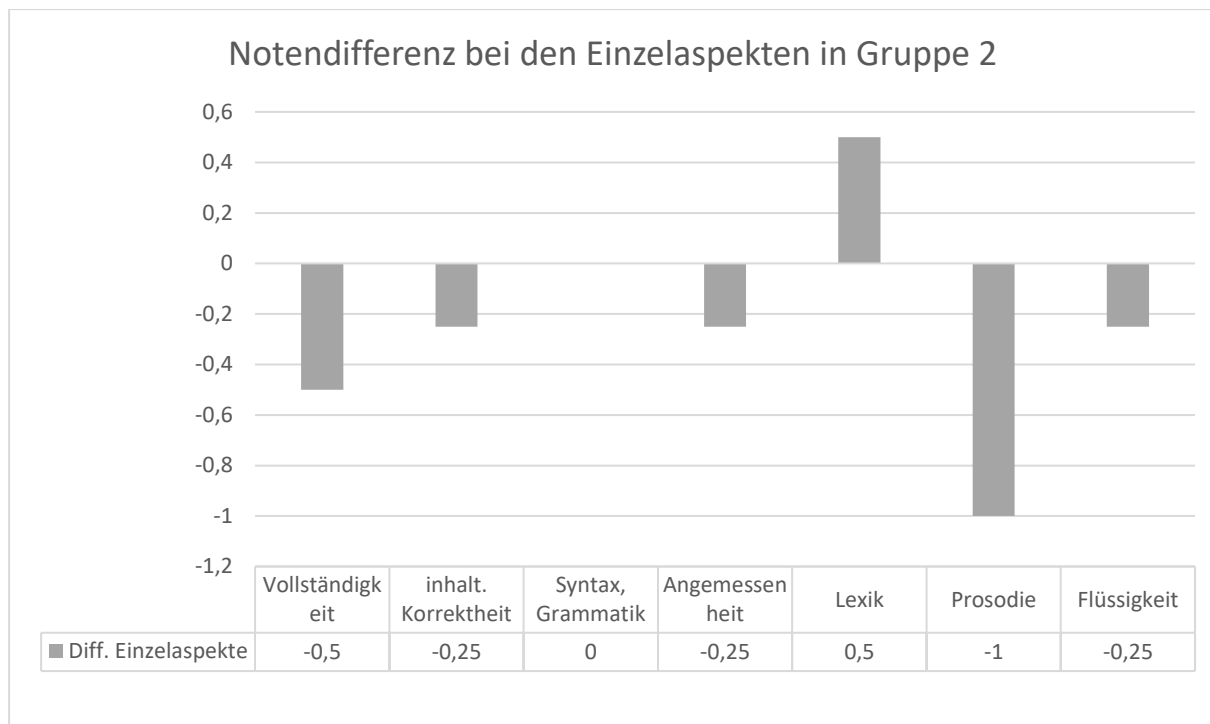


Abbildung 14: Differenz der Noten für die Einzelaspekte bei Gruppe 2 (eigene Erhebung)

Zusammenfassend kann an dieser Stelle festgehalten werden, dass im vorliegenden Versuch die Noten in beiden Gruppen vor allem bei den kommunikativ-sprachlichen Aspekten schwankten. In beiden Gruppen war „Prosodie“ durchschnittlich den größten Schwankungen unterlegen. Zudem schwankte bei beiden Teilnehmer_innen-Gruppen ein inhaltsbezogenes Kriterium vergleichsweise stärker, während eines im Durchschnitt beide Male gleich bewertet wurde. Es ist jedoch kein Muster zu erkennen, anhand dessen von tendenziellen Unterschieden zwischen den beiden nach Expertise eingeteilten Gruppen gesprochen werden könnte.

Betrachtet man die beiden Gruppen gemeinsam, ohne sie nach der vorgenommen Expertisedefinition zu trennen, zeichnet sich nochmals ein klareres Bild ab, nach dem die Notenvergabe bei den kommunikativ-sprachlichen Aspekten im Vergleich zu den inhaltsbezogenen Kriterien einer stärkeren Schwankung unterliegt (siehe Abbildung 15).

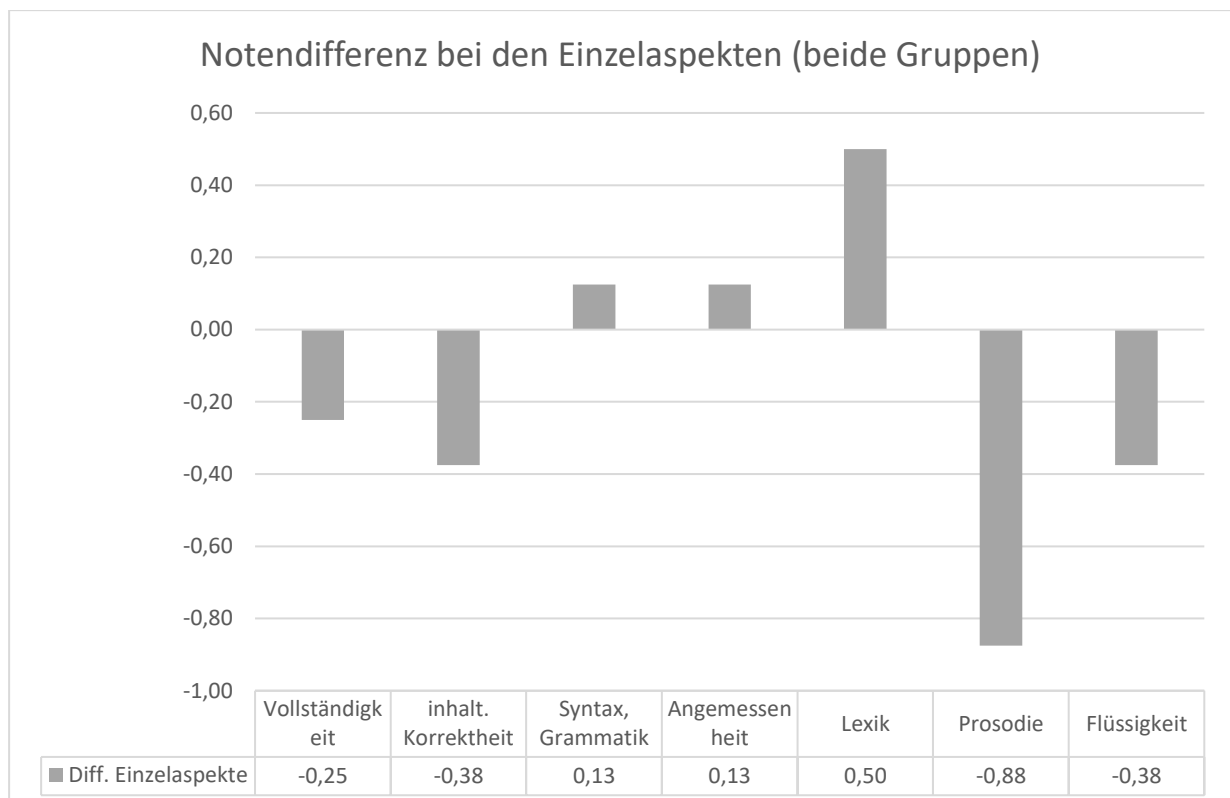


Abbildung 15: Differenz der Noten für die Einzelaspekte - beide Gruppen (eigene Erhebung)

In Abbildung 15 zeichnet sich eindeutig ab, dass die Benotung der „Lexik“ und „Prosodie“ insgesamt die größten Unterschiede zwischen den zwei Terminen aufwies. So sollte in einem letzten Schritt die Stimmung der vier bereits oben beschriebenen Personen, deren Stimmungswert die größten Schwankungen aufwies, ihren Beurteilungen dieser beiden Einzelaspekte gegenübergestellt werden. Abbildung 16 zeigt diese Gegenüberstellung für den Aspekt „Prosodie“.

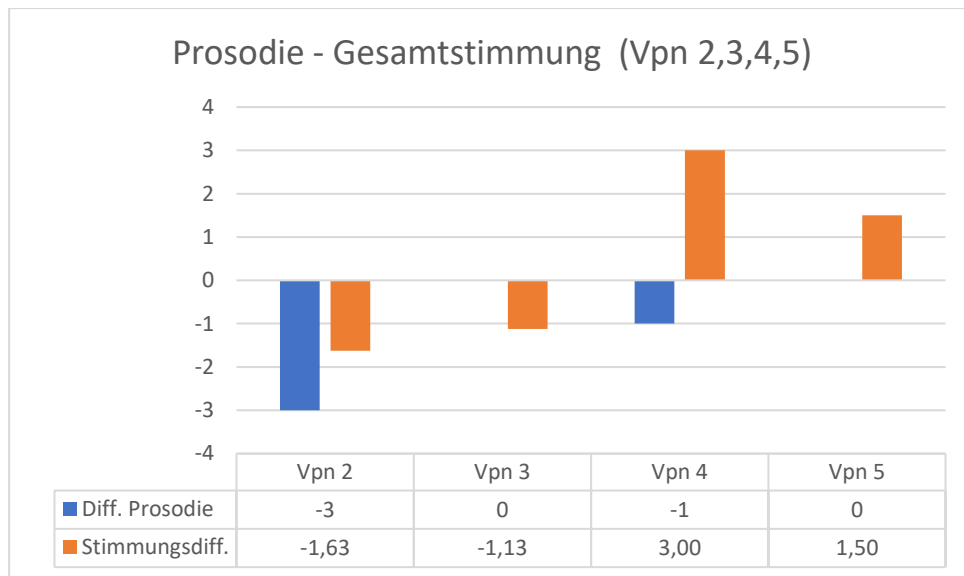


Abbildung 16: Gegenüberstellung Notendifferenz für „Prosodie“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)

Zwar fällt in diese Gruppe auch die Person hinein, bei der die Bewertung der „Prosodie“ die größte Differenz insgesamt aufwies (Vpn2), dennoch zeichnet sich hier kein klares Bild ab, das zu allgemeinen Schlussfolgerungen für eine tendenzielle Beeinflussung der Notenvergabe für „Prosodie“ durch die Stimmung in eine bestimmte Richtung führen würde. Dennoch weisen die zwei Personen mit der größten Stimmungsdifferenz auch Differenzen in der Beurteilung der Prosodie auf, auch wenn keine bestimmte Korrelation zu erkennen ist. Im Gegensatz hierzu ist bei der „Lexik“ eine gewisse Tendenz erkennbar. Abbildung 17 zeigt die Gegenüberstellung der Stimmungsdifferenz dieser vier Personen mit der Notenvergabe für diesen Einzelaspekt.

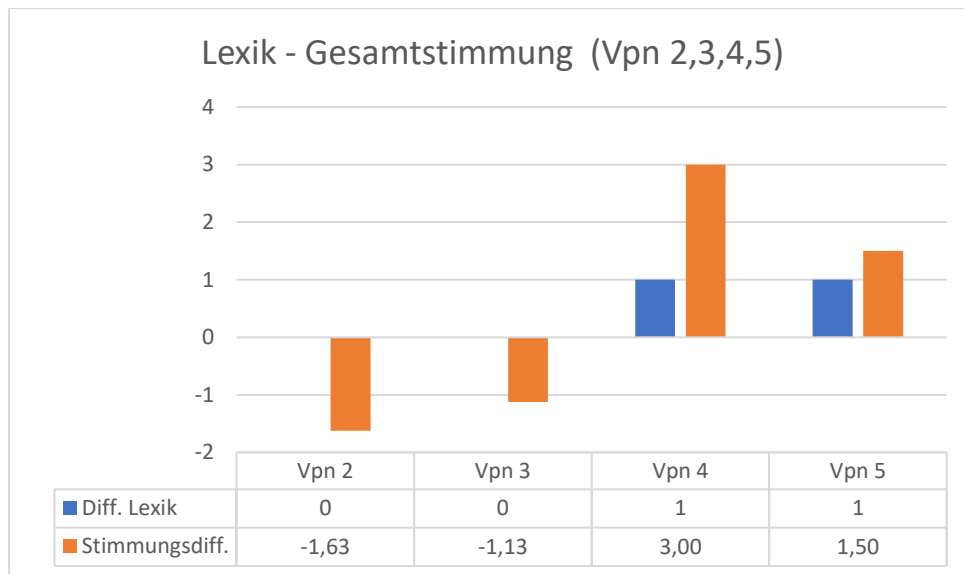


Abbildung 17: Gegenüberstellung Notendifferenz für „Lexik“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)

Es ist zu erkennen, dass bei zwei dieser Personen mit der Verbesserung der Stimmung auch eine bessere Note für die „Lexik“ vergeben wurde, während bei den anderen zwei Teilnehmer_innen eine Stimmungsverschlechterung keinen Einfluss auf die Notenvergabe bei diesem Einzelaspekt hatte. Ein Blick auf die Stimmungswerte dieser Personen (siehe Abbildung 4 in Kapitel weiter oben) zeigt, dass die Vpn 2 und 3 beim zweiten Termin eher schlechter Stimmung waren, während sie beim ersten Termin eher positiv gestimmt waren. Bei Vpn 4 und 5 war die Situation umgekehrt. Was hier erhoben wurde, würde mit der Annahme übereinstimmen, dass die zwei positiver gestimmten Personen beim zweiten Termin gemäß der FAIT eine heuristische Verarbeitungsweise annahmen, was sich in den besser vergebenen Noten für „Lexik“ abzeichnete. Die schlechter gelaunten Personen konzentrierten sich demnach womöglich auf die inhaltsbezogenen Parameter und weniger auf Aspekte wie die „Lexik“. Dieses Muster ist jedoch nicht in den Angaben zur „Prosodie“ (Abbildung 16) zu erkennen und zeichnet sich auch nicht bei den Notendifferenzen zu „Angemessenheit“ (siehe Abbildung 18) und „Flüssigkeit“ (Abbildung 19) ab.

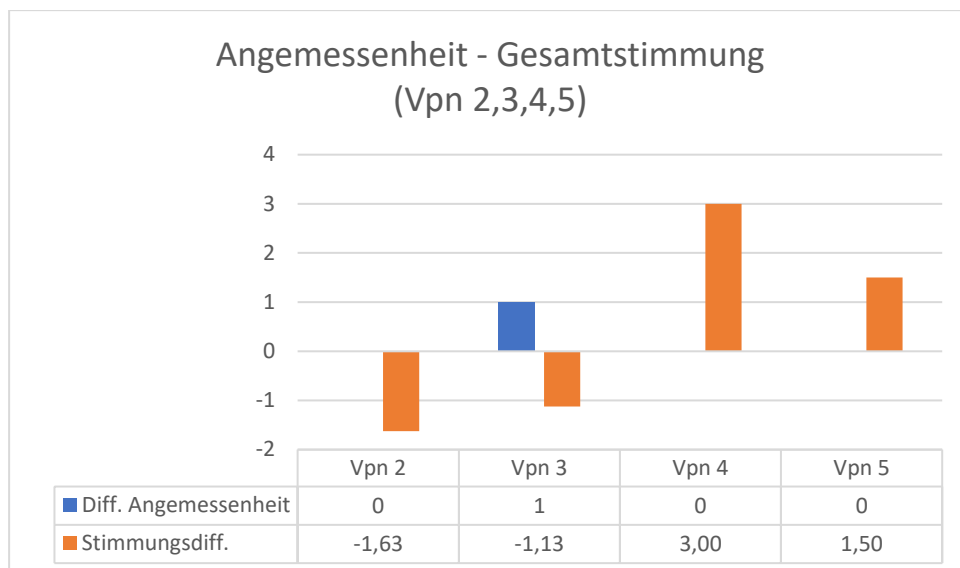


Abbildung 18: Gegenüberstellung Notendifferenz für „Angemessenheit“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)

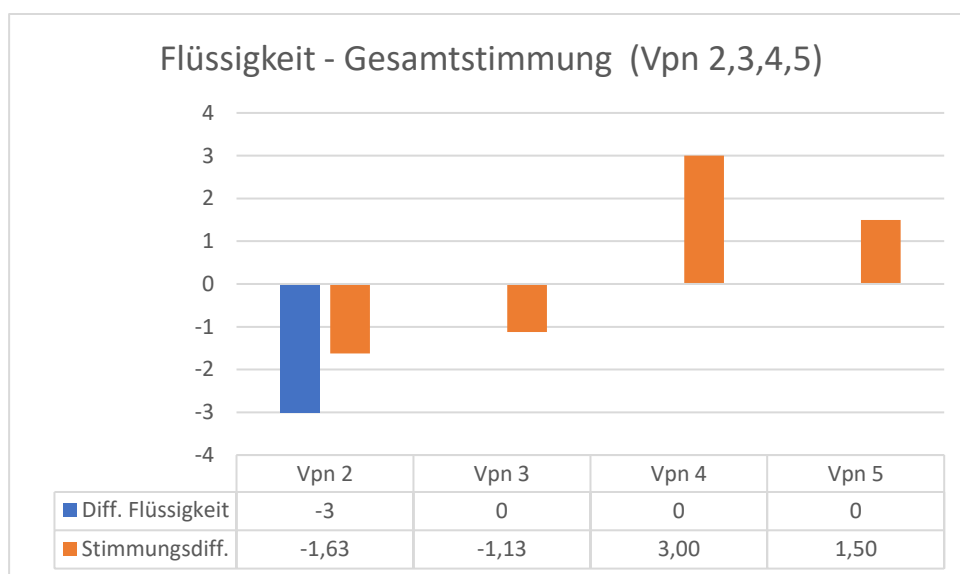


Abbildung 19: Gegenüberstellung Notendifferenz für „Flüssigkeit“ und der Stimmungsdifferenz bei Vpn 2, 3, 4 und 5 (eigene Erhebung)

Auch wenn diese Auswertung auf nur wenigen Daten beruht, könnte es hiervon ausgehend interessant sein zu untersuchen, ob die Stimmung bei der Bewertung von Verdolmetschungen womöglich nicht alle, sondern nur einige kommunikativ-sprachliche Einzelaspekte beeinflusst, wie beispielsweise die „Lexik“. Denn auch Behr (2013a) merkte an, dass erfahrene Dolmetscher_innen bestimmten Aspekten – in ihrer Untersuchung dem Aspekt Prosodie – eine

geringere Bedeutung zuweisen, da ihnen bewusst ist, dass diese Aspekte einen subjektiven Charakter aufweisen (Behr 2013a:304). Dennoch war im Gegensatz dazu, wie oben beschrieben, die Prosodie in diesem Versuch der Aspekt, der am meisten variierte, jedoch hier nicht mit der Stimmung zu korrelieren schien. Zusammenfassend kann an dieser Stelle festgehalten werden, dass zum einen weitere Untersuchungen notwendig sind, um den Stimmungseinfluss auf einzelne kommunikativ-sprachliche Aspekte beim Bewerten von Dolmetschungen zu untersuchen. Andererseits wäre es auch von Interesse, potenzielle andere Faktoren zu untersuchen, die einen Einfluss auf die Beurteilung von Einzelaspekten zu haben scheinen, jedoch scheinbar nicht mit der Stimmung zusammenhängen.

6. Diskussion

Das Ziel der vorliegenden Arbeit war es, den Einfluss von Stimmungen beim Bewerten von derselben Dolmetschaufnahme zu zwei verschiedenen Terminen zu untersuchen. Dabei handelte es sich um eine Untersuchung im didaktischen Setting, in dem Studierende im Rahmen eines Peerfeedbacks Dolmetschungen allgemein und in Hinblick auf ihre einzelnen Qualitätsparameter bewerten sollten. Dabei sollte auch der Aspekt der Expertise miteinbezogen werden. Um diesem Forschungsvorhaben nachzugehen, war zunächst eine theoretische Auseinandersetzung mit den hierfür relevanten Konzepten der Dolmetschqualität, der Expertise und der Stimmung notwendig. Die Arbeit veranschaulicht deutlich, wie variabel und relativ das soziale Konstrukt der Qualität ist. So hängt auch die Bewertung von vielen Faktoren ab. Dies reicht von der Situation bzw. dem Kontext bis hin zu der Person, die bewerten soll, und ihrer Perspektive. So kann in verschiedenen Bereichen bzw. sozialen Systemen eine Bewertung aus unterschiedlichen Blickwinkeln erfolgen. Eine Möglichkeit ist die Bewertung im didaktischen Bereich, wobei hier auch beispielsweise Studierende ihre Kolleg_innen bewerten können. Bei der Bewertung kommen zudem zahlreiche andere Faktoren hinzu, die in die Situation hineingetragen werden, wie Erwartungen, Normen, die Erfahrung oder das Befinden der Rezipient_innen. Um dem Aspekt der Erfahrung nachzugehen, der eng mit der Expertise verbunden ist, erfolgte daher auch ein Blick in die Expertiseforschung, die auch im Bereich des Dolmetschens ein wichtiges Forschungsgebiet darstellt.

Eine Auseinandersetzung mit dem Begriff des Affekts hat gezeigt, dass auch hier Begrifflichkeiten nicht immer eindeutig definiert werden können. Die vorliegende Arbeit interessierte sich für den Stimmungseinfluss, weshalb eine Betrachtung des Begriffs der Stimmung in Abgrenzung zu anderen Affektwörtern zentral war. Es wurde erläutert, dass die Stimmung Einfluss auf das Urteilen haben kann, was auch im Rahmen der FAIT postuliert wird.

Im empirischen Versuch, der im Rahmen dieser Arbeit durchgeführt wurde, sollte zunächst untersucht werden, welche Unterschiede sich bei der Bewertung derselben Simultandolmetschung, die unwissentlich zu zwei unterschiedlichen Terminen durch Studienkolleg_innen vorgenommen wurde, ergeben. Zudem sollte die Rolle der Stimmungen näher beleuchtet werden. Es konnte klar gezeigt werden, dass die Bewertung einer identischen Dolmetschaufnahme Unterschiede in der Notenvergabe zwischen zwei Terminen aufweist. Dabei bewertete die Mehrheit der Versuchspersonen die Dolmetschung zweimal anders. Auch unter den einzelnen Studierenden ergaben sich Unterschiede an beiden Terminen. So wurden beide Male Noten zwischen Sehr Gut (1) und Genügend (4) vergeben. Auf den ersten Teil der

gestellten Frage kann nach Auswertung der erhobenen Daten somit geantwortet werden, dass es sowohl innerhalb der beiden Versuchsgruppen als auch bei der Mehrheit der einzelnen Student_innen zu unterschiedlichen Noten zu den zwei Versuchsterminen kam. Dies zeigt eindeutig, wie variabel Qualitätsbewertung sein kann.

Die erste Forschungsfrage der vorliegenden Arbeit fragte weiters nach der Rolle der Stimmungen bei der Bewertung. Bei der Betrachtung der Stimmungen war zu sehen, dass an beiden Tagen Personen anwesend waren, die (eher) positiv gestimmt und (eher) negativ gestimmt waren. Außerdem unterschied sich die Stimmung der einzelnen Person jeweils zu den verschiedenen Versuchszeitpunkten. Dennoch konnte bei der gemeinsamen Betrachtung aller Versuchsteilnehmer_innen keine Tendenz erkannt werden. Auch die Betrachtung der vier Personen mit den größten Stimmungsdifferenzen, unabhängig von der ursprünglichen Gruppeneinteilung, ergab kein klares Bild. Somit kann für die Frage, welche Rolle die Stimmungen bei der Bewertung spielen, festgehalten werden, dass im vorliegenden Versuch keine aussagekräftige Korrelation gefunden wurde.

Für das Forschungsinteresse der vorliegenden Arbeit war zudem der Aspekt der Expertise zentral. So wurde weiters die Frage gestellt, welche Unterschiede sich bei der Bewertung aufgrund des Studienfortschritts der bewertenden Personen ergeben. Die Untersuchung der Gesamtevaluation der Studierenden nach ihrer Expertise in zwei Gruppen unterteilt, ergab in diesem empirischen Versuch jedoch keine eindeutige Tendenz und es zeigte sich bei der Vergabe der Noten für die untersuchte Aufnahme kein Zusammenhang mit der Expertise der Teilnehmer_innen. So konnten keine Muster in Hinblick auf die Bewertungsunterschiede erkannt werden, die sich aus dem Expertisegrad ergeben würden. Rundet man die durchschnittliche Gesamtbenotung aller Teilnehmer_innen, so ist zudem zu sehen, dass beide Gruppen im Durchschnitt zu beiden Zeitpunkten die gleiche Note vergaben. Bezieht man hier den Stimmungseinfluss mit ein, so konnte in Hinblick auf eine Korrelation zwischen Notenvergabe und Stimmung bei Gruppe 1 keine Tendenz aufgezeigt werden. In der zweiten Gruppe mit mehr Expertise zeichnete sich hingegen eine Korrelation ab. Mit der Veränderung der Stimmung ging in dieser Gruppe bei drei von vier Personen auch eine Veränderung der Note in die gleiche Richtung einher. Bei der Trennung nach den zwei Gefühlsdimensionen war erkennbar, dass die Dimension der Gestimmtheit hier eher eine Rolle spielte. An dieser Stelle wären daher weitere Untersuchungen mit mehr Personen notwendig, die sich auf die Dimension der Gestimmtheit konzentrieren und deren Einfluss auf die Notenvergabe erheben. In Hinblick auf die zweite Forschungsfrage kann somit zusammenfassend festgehalten werden, dass in Hinblick auf die Vergabe der Noten kein

Zusammenhang mit dem Studienfortschritt bzw. dem Grad der Expertise festgestellt werden konnte. Im Gegensatz dazu war bei der Einbeziehung der Stimmung eine Tendenz bei der Gruppe mit mehr Expertise zu erkennen, die darauf hindeutete, dass die Bewertungen eher mit den Stimmungen korrelierten. An dieser Stelle sind somit weitere Untersuchungen notwendig, um dem Stimmungseinfluss bei der Evaluation durch Studierende nachzugehen.

Schließlich war es auch Ziel der vorliegenden Arbeit, der Frage nachzugehen, inwieweit sich die Bewertungen der beiden Gruppen in Hinblick auf die inhaltsbezogenen bzw. formalen Qualitätsparameter unterscheiden. Bei der Auswertung der erhobenen Daten konnte in Hinblick auf die Evaluation der Einzelaspekte festgestellt werden, dass in beiden Gruppen die kommunikativ-sprachlichen Aspekte im Vergleich zu den inhaltlichen Parametern stärker schwankten. In beiden Versuchsgruppen schwankte der Qualitätsparameter „Prosodie“, der in die Kategorie der kommunikativ-sprachlichen Aspekte fällt, am stärksten. In Hinblick auf die inhaltsbezogenen Parameter war jeweils in beiden Studierendengruppen ein Aspekt zu finden, der beide Male im Durchschnitt die gleiche Bewertung erhalten hat. Eine Betrachtung nach dem Studienfortschritt zeigte jedoch keine Tendenz. In Bezug auf die letzte Forschungsfrage soll an dieser Stelle somit hervorgehoben werden, dass sich in Hinblick auf die Evaluation der einzelnen Qualitätsparameter kein Bild zeigte, das einen Zusammenhang zwischen der schwankenden Evaluation der Einzelaspekte und dem Expertisegrad vermuten ließe.

An dieser Stelle soll nochmals festgehalten werden, dass im Gegensatz zu den Erkenntnissen bei Behr (2013a) der Einzelaspekt „Prosodie“ am meisten schwankte. Um zu untersuchen, warum dieser Aspekt bei Bewertungen durch Studierende – oder möglicherweise bei Beurteilungen im didaktischen Bereich allgemein – schwankt, wären hier weitere Untersuchungen notwendig. Während bei der „Prosodie“ jedoch keine Tendenz erkennbar war, nach der die Stimmung die Bewertung dieses Aspektes in eine bestimmte Richtung lenken würde, konnte bei der Bewertung der „Lexik“ durch die Vpn 2,3,4 und 5 eine leichte Tendenz erkannt werden, die jedoch nicht bei den anderen Einzelaspekten erkennbar war. So wurde an dieser Stelle festgehalten, dass weitere Untersuchungen notwendig sind, um herauszufinden, ob womöglich einige kommunikativ-sprachliche Aspekte eher von Stimmungen beeinflusst werden als andere.

Zusammenfassend kann nochmals festgehalten werden, dass der vorliegende Versuch eindeutig gezeigt hat, dass eine Bewertung derselben Verdolmetschung durch dieselben Versuchsteilnehmer_innen, die an zwei verschiedenen Tagen erfolgt ist, anders ausfiel. Dabei wurden bei den kommunikativ-sprachlichen Aspekten mehr Schwankungen in der Notengebung beobachtet. Eine Trennung nach der Expertise, wie sie in diesem Versuch

vorgenommen wurde, hat kein klares Bild ergeben auch die Stimmung korrelierte in keiner klar erkennbaren Weise mit den Bewertungen. Dennoch konnten einige Punkte aufgezeigt werden, die für weiterführende Untersuchungen von Interesse sein könnten. Schließlich verdeutlichen die Ergebnisse, wie variabel und relativ Qualitätsbewertungen durch einzelne Personen sein können. Die Autorin hofft demnach, mit dieser Arbeit einen Beitrag zur Qualitätsforschung im Konferenzdolmetschen beigetragen zu haben.

7. Bibliografie

- AIIC (2019). JOIN AIIC. <https://aiic.org/site/world/join> (Stand: 27.06.2023).
- Alonso Bacigalupe, Luis (2013). Interpretation quality: from cognitive constraints to market limitations. In: Barranco-Droege, Rafael; Pradas Macías, E. Macarena & García Becerra, Olalla (Hg.) *Quality in interpreting: widening the scope. Volume 2*. Granada: Editorial Comares, 9-33.
- Barik, Henrik C. (1971). A Description of Various Types of Omissions, Additions and Errors of Translation Encountered in Simultaneous Interpretation. *Meta* 16 (4), 199–210.
- Bartłomiejczyk, Magdalena (2007). Interpreting Quality as Perceived by Trainee Interpreters. *The Interpreter and Translator Trainer* 1 (2), 247-267.
- Behr, Martina (2013a). *Evaluation und Stimmung. Ein neuer Blick auf Qualität im (Simultan-) Dolmetschen*. Berlin: Frank & Timme.
- Behr, Martina (2013b). Zur Einbeziehung kommunikationswissenschaftlicher Erkenntnisse bei der Evaluation von Dolmetschleistungen. In: Barranco-Droege, Rafael; Pradas Macías, E. Macarena & García Becerra, Olalla (Hg.) *Quality in interpreting: widening the scope. Volume 1*. Granada: Editorial Comares, 85-99.
- Behr, Martina (2020). *Dolmetschen: Komplexität, Methodik, Modellierung*. Berlin: Frank & Timme.
- Berger, Peter & Luckmann, Thomas (1969). *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. Frankfurt am Main: Fischer Verlag.
- Berger, Peter & Luckmann, Thomas (²⁵2013). *Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie*. Frankfurt am Main: Fischer Verlag.
- Briner, Rob B.; Parkinson, Brian; Reynolds, Shirley & Totterdell, Peter (Hg.) (2000). *Stimmungen: Struktur, Dynamik und Beeinflussungsmöglichkeiten eines psychologischen Phänomens*. Stuttgart: Klett-Cotta.
- Bühler, Hildegund (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* 5 (4), 231-235.
- Burr, Vivien (1995). *An Introduction to Social Constructionism*. London/New York: Routledge.
- Casado, Beatriz; Frauenfelder, Uli H.; Künzli, Alexander & Moser-Mercer, Barbara (2000). Searching to Define Expertise in Interpreting. In: Dimitrova, Birgitta Englund & Hyltenstam, Kenneth (Hg.) *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives*. Amsterdam/Philadelphia: John Benjamins, 107-131.

- Chabasse, Catherine (2009). *Gibt es eine Begabung für das Simultandolmetschen? Erstellung eines Dolmetscheignungstests mit Schwerpunkt Simultandolmetschen*. Berlin: SAXA.
- Cirillo, Letizia (2010). Managing affect in interpreter-mediated institutional talk: examples from the medical setting. *The Journal of Specialised Translation* 14, 55-79.
- Clore, Gerald L. & Schwarz, Norbert (1983). Mood, Misattribution, and Judgements of Well-Being: Informative and Directive Functions of Affective States. *Journal of Personality and Social Psychology* 45 (3), 513-523.
- Clore, Gerald L. & Schwarz, Norbert (2003). Mood as Information: 20 Years Later. *Psychological Inquiry* 14 (3&4), 296-303.
- Collados Aís, Ángela; Iglesias Fernández, Emilia; Pradas Macías, E. Macarena & Stévaux, Elisabeth (Hg.) (2011). *Qualitätsparameter beim Simultandolmetschen. Interdisziplinäre Perspektiven*. Tübingen: Narr Francke Attempto.
- Dastyar, Vorya (2019). *Dictionary of Education and Assessment in Translation and Interpreting Studies (TIS)*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Déjean Le Féal, Karla (1990). Some Thoughts on the Evaluation of Simultaneous Interpretation. In: Bowen, David & Bowen, Margareta (Hg.) *Interpreting: Yesterday, Today and Tomorrow*. New York: Binghamton, 154-160.
- Déjean Le Féal, Karla (1998). Didaktik des Dolmetschens. In: Hönig, Hans G.; Kußmaul, Paul; Schmitt, Peter A. & Snell-Hornby, Mary (Hg.) *Handbuch Translation*. Tübingen: Stauffenburg, 361-367.
- den Ouden, Barbara (2017). *Translation und Emotion: Untersuchung einer besonderen Komponente des Dolmetschens*. Berlin: Frank & Timme.
- Döring, Nicola (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Berlin/Heidelberg: Springer.
- Ericsson, K. Anders (2000). Expertise in interpreting. An expert-performance perspective. *Interpreting* 5 (2), 189-222.
- Eysenck, Michael W. & Keane, Mark T. (2020). *Cognitive Psychology. A Student's Handbook*. London/New York: Psychology Press.
- Fünfer, Sarah (2013). *Mensch oder Maschine? Dolmetscher und maschinelles Dolmetschsystem im Vergleich*. Berlin: Frank & Timme.
- Funke, Frederik (2006). Online- und Offlinevergleich visueller Analogskalen mit 4- und 8-stufig skalierten Likert-Skalen bei einem Fragebogen zum Verhalten in sozialen Gruppen. In: Rehberg, Karl-Siegbert (Hg.) *Soziale Ungleichheit, kulturelle*

- Unterschiede: Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München. Teilbd. 1 und 2.* Frankfurt am Main: Campus Verlag, 4826-4838.
- Funke, Frederik (2016). A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Science Computer Review* 34 (2), 244-254.
- Furmanek, Olgierda (2006). The role of emotions in establishing meaning: implications for interpreting. *Journal of Specialised Translation* 5, 57-71.
- Garzone, Giuliana (2003). Reliability of quality criteria evaluation in survey research. In: Collados Aís, Angela; Fernández Sánchez, M^a Manuela & Gile, Daniel (Hg.) *La evaluación de la calidad en interpretación: investigación*. Granada: Editorial Comares, 23-30.
- Gile, Daniel (2001). L'évaluation de la qualité de l'interprétation en cours de formation. *Meta* 46 (2), 379-393.
- Gile, Daniel (2003). Quality assessment in conference interpreting: methodological issues. In: Collados Aís, Angela; Fernández Sánchez, M^a Manuela & Gile, Daniel (Hg.) *La evaluación de la calidad en interpretación: investigación*. Granada: Editorial Comares, 109-123.
- Grbić, Nadja (2008). Constructing interpreting quality. *Interpreting* 10 (2), 232–257.
- Gruber, Hans & Mandl, Heinz (1992). Begabung und Expertise. In: Hany, Ernst A. & Nickel, Horst (Hg.) *Begabung und Hochbegabung. Theoretische Konzepte - Empirische Befunde - Praktische Konsequenzen*. Bern/Göttingen/Toronto/Seattle: Hans Huber, 59-73.
- Hacking, Ian (³2002). *Was heißt „soziale Konstruktion“? Zur Konjunktur einer Kampfvokabel in den Wissenschaften*. Frankfurt am Main: Fischer Verlag.
- Hale, Sandra & Napier, Jemina (2013). *Research Methods in Interpreting. A Practical Resource*. London/New York: Bloomsbury.
- Harvey, Lee & Green, Diana (1993). Defining quality. *Assessment and Evaluation in Higher Education* 18 (1), 9–34.
- Harvey, Lee & Green, Diana (2000). Qualität definieren. Fünf unterschiedliche Ansätze. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hg.) *Qualität und Qualitätssicherung im Bildungsbereich; Schule, Sozialpädagogik, Hochschule*. Basel/Weinheim: Beltz, 17-39.
- Heid, Helmut (2000). Qualität. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. In: Helmke, Andreas; Hornstein, Walter & Terhart, Ewald (Hg.)

- Qualität und Qualitätssicherung im Bildungsbereich; Schule, Sozialpädagogik, Hochschule.* Basel/Weinheim: Beltz, 41-51.
- Hejl, Peter M. (2003). Konstruktion der sozialen Konstruktion. Grundlinien einer konstruktivistischen Sozialtheorie. In: Gumin, Heinz & Meier, Heinrich (Hg.) *Einführung in den Konstruktivismus.* München/Zürich: Piper, 109-146.
- Herbert, Jean (1952). *The Interpreter's Handbook: How to become a conference interpreter.* Genève: Georg.
- Herbert, Jean (1978). How conference interpreting grew. In: Gerver, David & Sinaiko, H. Wallace (Hg.) *Language interpretation and communication.* New York: Plenum Press, 5-10.
- Hoffmann, Robert R. (1997). The cognitive psychology of expertise and the domain of interpreting. *Interpreting* 2 (1-2), 189-230.
- Hubscher-Davidson, Séverine (2018). *Translation and Emotion. A Psychological Perspective.* New York: Routledge.
- ISO (2015). Quality management systems – Fundamentals and vocabulary. <https://www.iso.org/obp/ui/en/#iso:std:iso:9000:ed-4:v1:en> (Stand: 28.11.2023).
- Kahane, Eduardo (2000). Thoughts on the quality of interpretation. https://aiic.org/document/364/AIICWebzine_May2000_1_KAHANE_Thoughts_on_the_quality_of_interpretation_EN.pdf (Stand 09.08.2023).
- Kalina, Sylvia (1998). *Strategische Prozesse beim Dolmetschen. Theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen.* Tübingen: Gunter Narr Verlag.
- Kalina, Sylvia (2002). Fragestellungen der Dolmetschwissenschaft. In: Best, Joanna & Kalina, Sylvia (Hg.) *Übersetzen und Dolmetschen. Eine Orientierungshilfe.* Basel/Tübingen: Francke, 30-43.
- Kalina, Sylvia (2004). Zum Qualitätsbegriff beim Dolmetschen. *Lebende Sprachen* 49 (1), 2-8.
- Kalina, Sylvia (2005). Quality Assurance for Interpreting Processes. *Meta* 50 (2), 768-784.
- Koskinen, Kaisa (2020). *Translation and Affect. Essay on sticky affects and translational affective labour.* Amsterdam/Philadelphia: John Benjamins.
- Kurz, Ingrid (1989). Conference Interpreting – User Expectations. In: Lindberg Hammond, Deanna (Hg.) *Coming of Age.* Medford N.J.: Learned Information Inc., 143-147.
- Kurz, Ingrid (2003). Quality from the user perspective. In: Collados Aís, Angela; Fernández Sánchez, Manuela & Gile, Daniel (Hg.) *La evaluación de la calidad en interpretación: investigación.* Granada: Editorial Comares, 3-22.

- Kutz, Wladimir (2005). Zur Bewertung der Dolmetschqualität in der Ausbildung von Konferenzdolmetschern. *Lebende Sprachen* 1, 14-34.
- Liu, Minhua (2011). Methodology in interpreting studies. A methodological review of evidence-based research. In: Nicodemus, Brenda & Swabey, Laurie (Hg.) *Advances in Interpreting Research: Inquiry in action*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 85-119.
- Luhmann, Niklas (1991). *Soziologie des Risikos*. Berlin: de Gruyter.
- Mack, Gabriele (2002). Die Beurteilung professioneller Dolmetschleistungen. In: Best, Joanna & Kalina, Sylvia (Hg.) *Übersetzen und Dolmetschen. Eine Orientierungshilfe*. Basel/Tübingen: Francke, 110-119.
- Martí, Mabel Abril & Martin, Anne (2003). Teaching interpreting: some considerations on assessment. In: Collados Aís, Angela; Fernández Sánchez, M.^a Manuela; Pradas Macías, E. Macarena; Sánchez Adam, Concepción & Stévaux, Elisabeth (Hg.) *La evaluación de la calidad en interpretación: docencia y profesión*. Granada: Editorial Comares, 197-207.
- Mendelsohn, Gerald A.; Russell, James A. & Weiss, Anna (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology* 57 (3), 493-502.
- Moser-Mercer, Barbara (1996). Quality in interpreting: some methodological issues. *The Interpreters' Newsletter* 7, 43-55.
- Moser-Mercer, Barbara (2000). The rocky road to expertise in interpreting: Eliciting knowledge from learners. In: Kadric, Mira; Kaindl, Klaus & Pöchhacker, Franz (Hg.) *Translationswissenschaft. Festschrift für Mary Snell-Hornby zum 60. Geburtstag*. Tübingen: Stauffenburg, 339-352.
- Pöchhacker, Franz (1994). *Simultandolmetschen als komplexes Handeln*. Tübingen: Gunter Narr.
- Pöchhacker, Franz (1998). Situative Zusammenhänge. In: Hönig, Hans G.; Kußmaul, Paul; Schmitt, Peter A. & Snell-Hornby, Mary (Hg.) *Handbuch Translation*. Tübingen: Stauffenburg, 327-330.
- Pöchhacker, Franz (2004). *Introducing Interpreting Studies*. London/New York: Routledge.
- Pöchhacker, Franz (2012). Qualität, die man versteht: ein funktional-kognitiver Ansatz. In: Ahrens, Barbara; Albl-Mikasa, Michaela & Sasse, Claudia (Hg.) *Dolmetschqualität in Praxis, Lehre und Forschung. Festschrift für Sylvia Kalina*. Tübingen: Narr Francke Attempto, 19-32.

- Pöchhacker, Franz (2013). Researching quality: A two-pronged approach. In: Barranco-Droege, Rafael; Pradas Macías, E. Macarena & García Becerra, Olalla (Hg.) *Quality in interpreting: widening the scope. Volume I*. Granada: Editorial Comares, 33-55.
- Pradas Macías, E. Macarena & Zwischenberger, Cornelia (2022). Quality and norms in conference interpreting. In: Albl-Mikasa, Michaela & Tiselius, Elisabet (Hg.) *The Routledge Handbook of Conference Interpreting*. Abingdon/New York: Routledge, 234-257.
- Prunč, Erich (³2012). *Entwicklungslinien der Translationswissenschaft. Von den Asymmetrien der Sprachen zu den Asymmetrien der Macht*. Berlin: Frank & Timme.
- Riccardi, Alessandra (2002). Evaluation in interpretation. Macrocriteria and microcriteria. In: Hung, Eva (Hg.) *Teaching Translation and Interpreting 4. Building Bridges*. Amsterdam/Philadelphia: John Benjamins, 115-126.
- Riccardi, Alessandra (2003). The relevance of interpreting strategies for defining quality in interpreting. In: Collados Aís, Angela; Fernández Sánchez, M^a Manuela & Gile, Daniel (Hg.) *La evaluación de la calidad en interpretación: investigación*. Granada: Editorial Comares, 257-265.
- Riccardi, Alessandra (2005). Qualitätskriterien beim Simultandolmetschen. In: Zybatow, Lew N. (Hg.) *Translatologie – neue Ideen und Ansätze. Innsbrucker Ringvorlesung zur Translationswissenschaft IV*. Frankfurt am Main: Peter Lang, 99-112.
- Riccardi, Alessandra (2007). Mikrokriterien und Makrokriterien und ihre Rolle bei der Evaluierung von Dolmetschleistungen. In: Baumgarten, Nicole & House, Juliane (Hg.) *Translationskritik: Modelle und Methoden*. Bochum: AKS-Verlag Bochum, 187-205.
- Richter, Rudolf (²2016). *Soziologische Paradigmen. Eine Einführung in klassische und moderne Konzepte*. Wien: Facultas.
- Russell, James A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology* 39 (6), 1161-1178.
- Sawyer, David B. (2004). *Fundamental Aspects of Interpreter Education. Curriculum and Assessment*. Amsterdam/Philadelphia: John Benjamins.
- Scherer, Klaus (⁴2002). Emotion. In: Hewstone, M.; Jonas, K. & Stroebe W. (Hg.) *Sozialpsychologie. Eine Einführung*. Berlin: Springer, 165-213.
- Schneider, Wolfgang (1992). Erwerb von Expertise. Zur Relevanz kognitiver und nichtkognitiver Voraussetzungen. In: Hany, Ernst A. & Nickel, Horst (Hg.) *Begabung und Hochbegabung. Theoretische Konzepte - Empirische Befunde - Praktische Konsequenzen*. Göttingen: Hans Huber, 105-122.

- Schwarz, Norbert (2010). *Feelings-as-Information-Theory*.
http://sitemaker.umich.edu/norbert.schwarz/files/schwarz_feelings-as-information_7jan10.pdf (Stand: 12.08.2011).
- Schwarz, Norbert (2011). Feelings-as-Information Theory. In: Higgins, E. Tory; Kruglanski, Arie W. & Van Lange, Paul (Hg.) *Handbook of Theories of Social Psychology: Volume One*. London/New Delhi/Singapore/Thousand Oaks: SAGE Publications, 289-308.
- Seleskovitch, Danica (1976). Interpretation, a psychological approach to translation. In: Brislin, Richard W. (Hg.) *Translation: Applications and Research*. New York: Gardner Press, 92-116.
- Seleskovitch, Danica (1986). Comment: Who should assess an interpreter's performance? *Multilingua* 5 (4), 236.
- Setton, Robin (2007). Staying relevant: Interpreting in the information age. In: Jakobsen, Arnt Lykke; Mees, Inger M. & Pöchhacker, Franz (Hg.) *Interpreting Studies and Beyond. A Tribute to Miriam Shlesinger*. Copenhagen: Samfundslitteratur Press, 53-72.
- Shlesinger, Miriam (1997). Quality in simultaneous interpreting. In: Gambier, Yves; Gile, Daniel & Taylor, Christopher (Hg.) *Conference interpreting. Current trends in research*. Amsterdam/Philadelphia: John Benjamins, 123-131.
- Sunnari, Marianna (2003). Expert and novice performance in simultaneous interpreting: implications for quality assessment. In: Collados Aís, Angela; Fernández Sánchez, M^a Manuela & Gile, Daniel (Hg.) *La evaluación de la calidad en interpretación: investigación*. Granada: Editorial Comares, 235-247.
- Viezzi, Maurizio (1996). *Aspetti della Qualità in Interpretazione*. Trieste: SSLMIT.
- Viezzi, Maurizio (2007). Interpretation quality assessment: from theory to practice. In: Baumgarten, Nicole & House, Juliane (Hg.) *Translationskritik: Modelle und Methoden*. Bochum: AKS-Verlag Bochum, 171-186.
- Wassmann, Claudia (2017). Forgotten Origins, Occluded Meanings: Translation of Emotion Terms. *Emotion Review* 9 (2), 163-171.
- Winter, Miriam (2011). Expertise im Dolmetschen. In: Andres, Dörte & Behr, Martina (Hg.) *Interpretes Mundi – Deuter der Welt*. München: Martin Meidenbauer, 203-226.
- Wundt, Wilhelm (1897). *Outlines of psychology*. Leipzig: Wilhelm Engelmann.
- YouTube (o.J.). The Human-Animal Bond | Susan Little | TEDxOStateU.
<https://www.youtube.com/watch?v=jjxAewIBYJk> (Stand: 03.09.2023).

- ZTW (2005). Curriculum für das Masterstudium Translation. https://transvienna.univie.ac.at/fileadmin/user_upload/z_translationswiss/Studium/Curricula/Curriculum_Masterstudium_Translation_2015.pdf (Stand 12.11.2013).
- ZTW (o.J.). BEURTEILUNGSKRITERIEN SIMULTANDOLMETSCHEN. https://transvienna.univie.ac.at/fileadmin/user_upload/z_translationswiss/Studium/Masterstudium/BEURTEILUNGSKRITERIEN_SIM-Studierende.pdf (Stand: 17.07.2022).
- Zwischenberger, Cornelia (2013). *Qualität und Rollenbilder beim simultanen Konferenzdolmetschen*. Berlin: Frank & Timme.

Anhang

Anhang A: Anleitung zum Ausfüllen der Fragebögen

Anleitung zum Ausfüllen der Fragebögen

- Du hörst nun drei Aufnahmen von Verdolmetschungen aus dem Englischen ins Deutsche. Bitte höre sie mit dem Ziel an, diese anschließend zu bewerten.
- Gehe beim Anhören so vor, wie du es aus den Lehrveranstaltungen gewohnt bist. Parallel zur Verdolmetschung wird die Originalrede eingespielt. Die Dolmetschung und das Originalvideo sind auf separaten Audiokanälen auf jeweils einer Seite der Kopfhörer zu hören. Stelle alles nach deinen Bedürfnissen ein. Auch ein Blatt Papier für etwaige Notizen steht dir zur Verfügung.
- Fülle direkt nach jeder Aufnahme einen der beigefügten Bewertungsfragebögen aus. Gib immer zuerst **spontan** eine **Gesamtwertung** ab (Benotung von 1 = „Sehr Gut“ bis 5 = „Nicht Genügend“). Bewerte **anschließend** die **Einzelaspekte** und setze hierfür ein Kreuz im entsprechenden Kästchen (von 1 = „Sehr Gut“ bis 5 = „Nicht Genügend“).
- Nach der Bewertung **ALLER** Dolmetschaufzeichnungen hast du die Möglichkeit, die spontane Gesamtbeurteilung gegebenenfalls abzuändern. Wenn du dies tun möchtest, tue dies bitte im Punkt „Anmerkungen“.
- Am Ende des Versuchs, nach Anhören und Bewertung aller Aufnahmen, wirst du gebeten, einen Fragebogen zu persönlichen und studiumsspezifischen Angaben auszufüllen.

VIELEN DANK FÜR DIE TEILNAHME!

Anhang B: Bewertungsbogen zur Evaluation der Dolmetschungen

Bewertung der Aufnahme Nr. _____

Vpn: _____

Gesamtbeurteilung (Note von 1 bis 5): _____

INHALT

Vollständigkeit

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

Inhaltliche Korrektheit

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

Angemessenheit fürs Zielpublikum

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

ZIELSPRACHE

Syntax, Grammatik

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

Lexik (lexikalische/stilistische Aspekte)

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

DOLMETSCHFERTIGKEITEN & PRÄSENTATION

Prosodie

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

Flüssigkeit

	1	2	3	4	5	
Sehr Gut	☐	☐	☐	☐	☐	Nicht Genügend

Anmerkung: Nach Anhören aller Dolmetschungen möchte ich die Gesamtbeurteilung auf folgende Note abändern: _____

Anhang C: Fragebogen zu persönlichen und studiumsspezifischen Angaben

Persönliche und studiumsspezifische Angaben

Vpn: _____

Bitte trage die jeweiligen Informationen in die vorgesehenen Zeilen ein.

1. Schwerpunkt: _____

2. Sprachkombination: A-Sprache: _____

B-Sprache: _____

C-Sprache/n: _____ / _____ / _____

3. Anzahl absolvierter Semester: _____ Semester

4. Modulprüfung:

Hast du die Modulprüfung bereits vollständig absolviert? _____ (ja/nein)

Falls nein: Wann planst du, das erste Mal anzutreten? _____

Weitere Kommentare bzw. Informationen, die aus deiner Sicht relevant sind:

Neuste Erkenntnisse aus der Umfrageforschung weisen darauf hin, dass auch Informationen über das Allgemeinbefinden der teilnehmenden Personen in einem Versuch erhoben werden sollen.

Fülle daher bitte auch den beigegefügt Zusatzfragebogen aus und kreuze entsprechend an.

Anhang D: Fragebogen zur Erhebung der Stimmung

Zusatzfragebogen

Vpn: _____

→ Ich fühle mich in den letzten ca. 5 Tagen überwiegend...

	0	1	2	3	4	5	6	7	8	9	10	
sehr gestresst	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr erholt
	0	1	2	3	4	5	6	7	8	9	10	
sehr gut gelaunt	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr schlecht gelaunt
	0	1	2	3	4	5	6	7	8	9	10	
sehr müde	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	voller Energie
	0	1	2	3	4	5	6	7	8	9	10	
sehr beschwingt	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr niedergeschlagen

→ Ich fühle mich allgemein am heutigen Tag...

	0	1	2	3	4	5	6	7	8	9	10	
sehr wohl	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr unwohl
	0	1	2	3	4	5	6	7	8	9	10	
sehr träge	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	hellwach
	0	1	2	3	4	5	6	7	8	9	10	
sehr ausgeglichen	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr angespannt
	0	1	2	3	4	5	6	7	8	9	10	
sehr neugierig	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	sehr gelangweilt

VIELEN DANK FÜR DIE TEILNAHME!

Abstract (Deutsch)

Das Forschungsinteresse der vorliegenden Arbeit ist im didaktischen Setting zu verorten und befasst sich mit der Dolmetschqualität und ihrer Bewertung, wobei der Einfluss von Stimmungen und die Rolle der Expertise beim Evaluieren genauer untersucht wurde. Ausgehend von Qualität als soziales Konstrukt wurde die Variabilität der Beurteilung von Simultandolmetschungen betrachtet, die neben dem situativen Kontext auch von der Bewertungsperspektive abhängt. Dem folgte auch eine Auseinandersetzung mit dem Konzept der Expertise und eine Abgrenzung des Begriffs der Stimmung im Gegensatz zur Emotion.

Ausgehend von der FAIT nach Schwarz (2011) und der Arbeit von Behr (2013a) wurde ein empirischer Versuch geplant und am Zentrum für Translationswissenschaft in Wien durchgeführt. Dabei wurde untersucht, welche Unterschiede die Beurteilung durch Studierende von der gleichen Dolmetschaufzeichnung aufweisen und welche Rolle hierbei die Stimmungen spielen. Weiters wurde untersucht, ob sich die Beurteilungen aufgrund der Expertise unterscheiden. Schließlich wurde auch die Benotung einzelner Qualitätsparameter näher beleuchtet.

Die Auswertung der erhobenen Daten ergab, dass die Bewertungen der gleichen Verdolmetschung an zwei verschiedenen Versuchsterminen unterschiedlich ausgefallen sind. Dabei wurden im Vergleich zu den inhaltlichen Merkmalen mehr Schwankungen bei der Benotung der kommunikativ-sprachlichen Qualitätsparameter aufgezeigt, wobei vor allem der Aspekt „Prosodie“ betroffen war. Jedoch konnten bei der Betrachtung eines potenziellen Einflusses durch die Stimmungen und die Expertise keine klaren Korrelationen erkannt werden.

Stichwörter: Dolmetschqualität, Qualitätsbewertung, Qualitätskriterien, Expertise, Stimmungseinfluss

Abstract (Englisch)

The research interest of this thesis is dedicated to the didactic setting and deals with quality in interpreting and its assessment, whereby the influence of moods and the role of expertise in evaluation were examined in detail. Based on the notion of quality as a social construct, the variable nature of assessment in simultaneous interpreting was investigated. Quality assessment or evaluation depends not only on the situational context but also on the perspective. Further, the concepts of expertise and mood, as opposed to emotion, was looked at more closely.

Based on the FAIT by Schwarz (2011) and the study of Behr (2013a), an empirical quasi-experiment was conducted at the Centre for Translation Studies in Vienna. The aim was to investigate the differences between students' assessments of the same interpreting recording and the impact of moods. Further, potential differences related to different levels of expertise were examined. Finally, the assessment of individual quality parameters was analysed in more detail.

The data showed that the assessments of the same interpretation were different on two different days of testing. As for the assessment of the quality parameters, more variation was found in the communicative-linguistic criteria in comparison to content-related parameters, with the aspect of "prosody" being particularly affected. However, no clear correlations could be identified when considering the potential influence of moods and expertise.

Keywords: interpreting quality, quality assessment, quality criteria, expertise, influence of mood